Proceedings of LFG15

Miriam Butt and Tracy Holloway King (Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Contents

1 Editors' Note

We are pleased to bring you the 20th edition of the LFG Proceedings this year. Year after year, it has been a pleasure to read and comment on all of the papers in the proceedings and then publish them in this freely accessible on-line format. We would like to thank Dikran Karagueuzian at CSLI Publications for supporting and pioneering open access long before it became so intensely discussed in the wider publishing community.

The program committee for LFG15 were John Lowe and Ida Toivonen. We would like to thank them for coordinating the review process and liaising with the local organizer Ryo Otoguro in order to put together the final program that gave rise to this collection of papers. Thanks also go to the executive committee and the abstract and final paper reviewers, without whom the conference and the proceedings would not have been possible in this form.

The table of contents lists all the papers presented at the conference. Some papers were not submitted to the proceedings. For these papers, we suggest contacting the authors directly.

Hard Copy: All of the papers submitted to the LFG15 proceedings are available in one large pdf file, to be viewed and printed with Adobe Acrobat. The proceedings' file was created via pdflatex tools, for which we continue to be indebted to Stefan Müller. We thank Emma Pease at CSLI Publications for having accompanied the LFG Proceedings over the years and making sure they become accessible and stay accessible.

CONSTRUCTED MIDDLES IN MARORI: AN LFG ANALYSIS

I Wayan Arka

Australian National University
Udayana University

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King (Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

This paper discusses middles in Marori (Isolate, TNG; Indonesian Papua), contributing to the theoretical debate as to the best approach to middles. Marori data on middles shows that the valence and transitivity of a middle structure is constructed in both morphology and syntax. However, certain lexical properties of the predicate (e.g. inherent reflexivity) are important. It is demonstrated that LFG's parallel structure model is well suited to handle the properties of middles in Marori. Drawing insights from earlier LFG works on reflexives/reciprocals (Alsina 1996, Dalrymple et al. 1998, Rákosi 2008, Hurst 2010), a lexical-constructional analysis in LFG to account for the interface of morphology-syntax-semantics of middle expressions in Marori is proposed.

1 Introduction

The middle is part of a 'broad semantic-pragmatic domain including traditional voice categories and also semantic categories of transitivity and intransitivity' (Kemmer 1993).[*] The middle is conceived as a compromise voice category displaying characteristics of both the active (i.e. notionally from the dynamic Agent-oriented standpoint) and the passive (i.e. from the nondynamic Patient-oriented perspective) (Klaiman 1991:3). Important cross-linguistic characteristics of middle constructions include, among other things, their intransitive syntax with meanings typically associated with 'reflexivity', 'reciprocity', and 'passivity' (i.e. where the subject/agent is also affected). In English middles like *These kinds of walls paint easily*, for example, the verb is 'active' but the syntax is intransitive with subject being patientive. Unlike English (which lacks a middle marker), other languages may have a specific marker for this, e.g. *se* in Spanish, or *ma-* in Balinese. Marori is unusual in that the middle is constructed; its marking is, as we see below, parasitic to the verbal agreement morphology.

Verbs typically in middle constructions are verbs of grooming or self-indulgence. The same is true for Marori. Consider the grooming verb of 'combing' given in (1).[1] As seen, the marking of the reflexive meaning

[*] I gratefully acknowledge the support of ARC Discovery Grant (DP10100307). For helpful discussion, I thank the anonymous reviewers and audience at the LFG2015, in particular Andrew Spencer, John Payne and Miriam Butt.

[1] Abbreviations, alphabetically ordered: 1,2,3 (first, second and third person), A (Actor), AUX (auxiliary), CPLT (completive), D.AUX (dynamic auxiliary), DEIC (deictic), DUR (durative), F (feminine), FUT (future), HAB (habitual), LOC (locative), IRR (irrealis), MID (middle), NPL (nonplural), M (masculine), MP (macro present), NrPST (near past), O (object), P (Patient), PL (plural), POSS (possessive), PRES (present), Q (question marker), REAL (realis), RmPST (Remote Past), S (intransitive subject), STAT (stative), SG (singular), U (undergoer).

in the grooming verb is parasitic to the existing transitive morphology in this language. The prefix *i-* '1SG' is the Undergoer (U)/object prefix and the suffix *-du* '1SG' is the Actor (A)/subject suffix, encountered in highly transitive structures; see Tables 1 and 2.

(1) na pu tpab i-ngg-ra-du.
1SG hair comb 1SG.U-AUX-DUR-1SG.A.PRES
'I am combing my own hair'
(Lit. 'I am doing (my own) hair combing.')

The middle construction shown in (1) is one type, called Middle Type 1 or MID-1 for short. Marori has another type of middle construction, called Middle Type 2 (MID-2), featuring the invariant prefix *n-* in the U prefix slot. This is exemplified in (2).

(2) pake=na tifa=n-ngg-ra-mon.
there=1SG hide=MID-AUX-1SG.A.DUR.NrPST
'I was hiding (myself) there.'

Constructions shown in (1) and (2) are middles as they meet cross-linguistic and language-specific properties of middles, distinct from main voice types such as the active, or other constructions such as reflexives (as Marori does have a reflexive construction, see (15).

(3) John na=i tirfa=ri-ngg-ra-m.
John 1SG=U hide=1SG.U-AUX-DUR-3SG.A.NrPST
'John hid me.'

The issues raised by middle constructions in Marori include the following. Firstly, what governs the distribution of the two middle types? Secondly, to what extent can LFG capture constraints associated with the two types? In particular, of great interest is the fact that MID-1 shows up with the verb *abon* 'steal', as seen in (4).

(4) na=i (bosik=i) abon yu-nggo-ru.
1SG=U pig=U steal 1SG.U-AUX-1SG.A.FUT
'I will steal pigs.'

Sentence (4) has two intriguing properties. One property is the morphology-syntax transitivity mismatch. Note that the verb is morphologically intransitive. That is, it is in middle form with co-referential prefix and suffix. Hence the middle verb is an intransitive form, parasitically constructed through the transitive morphology. Syntactically, however, the verb is ambitransitive as it can appear in transitive syntax as seen by the possibility of it taking the object 'the pig'.

The other, seemingly puzzling property is that the predicate *abon* 'steal' can have its subject flagged by the undergoer clitic *=i* discussed later in Section 3. Note that *=i* is not possible with agentive non-middle predicates, either transitive or intransitive, as shown in (5). The ungrammaticality of the flagging with *=i* is indicated by putting the star (*) inside the brackets, i.e. (**=i*).

(5) a. Albert(*=i) na=i turpar**=i**-ngg-ra.
Albert=U 1SG=U kick=1SG.U-AUX-3.A.DUR.PRES
'Albert (often) kicks me'

b. na(*=i)fis kund-ra-mon.
1SG(=U) yesterday run-DUR-1NPL.A.NrPST
'I was running yesterday.'

The morphosyntax of Marori middles with the peculiarities briefly illustrated above is further discussed in a precise way within LFG in the ensuing sub-sections. The paper is organised as follows. In section 2, an overview of Marori morphosyntax is given. In section 3, the proposed analysis in LFG is outlined, consisting of the specifications of lexical entries and the formulation of Paradigm Function (PF) rules incorporating mechanism from (G)PFM (Generalised Paradigm Function Morphology) (Stump 2001, Spencer 2010, 2013) to handle Morphology-Syntax-semantics interface issues in middle constructions. In the last section, final remarks are provided regarding the implication of the present study.

2 Types of Marori Middles and their morphosyntax

2.1 Marori clausal structure

Marori is a non-configurational verb-final language, depicted in (6). The verbal predicative complex typically consists of a lexical predicate (X), immediately followed by a light/auxiliary verb (V), inflected for tense, aspect and mood (TAM). Certain lexical items of high frequency such as 'run' and 'walk' are directly affixed with TAM morphology. This TAM morphology is also for subject agreement. The TAM (subject) suffix sets are classified into two on the basis of aspectual properties; see Tables 1-2. Since middle marking makes use of these suffix sets, TAM meaning is also relevant to the discussion of middles in Marori (cf. section 3).

(6) NP*(=*i*) , X V
ARG(=U) LEXICAL PRED (inflected)

	(1a) IRR/FUT			(1b) NrPST (Completive)			(1c) RmPST (Completive)		
	1	2	3	1	2	3	1	2	3
SG	*-ru*	*-Ø*	*-Ø*	*-ben*	*-f*	*-f*	*-fori*	*-fi*	*-fi*
DU	*-ren*	*n--Ø*	*-Ø*	*-ben*	*n- -f*	*-f*	*-fori*	*n- -fi*	*-fi*
PL	*-men*	*n-(ri)m*	*-(ri)m*	*-freben* *-frendu*	*n- -(fre)f*	*(fre)f*	*-mbrofori*	*-mbrofi*	*mbrofi*

Table 1: Class 1 Argument suffixes in Marori

	(2a) REAL/MacroPRES (Completive/extended)			(2b) NrPST (Durative)			(2c) RmPST (Durative)		
	1	2	3	1	2	3	1	2	3
SG	*-du*	*-Ø*	*-Ø*	*-men*	*-m*	*-m*	*-maf*	*-maf*	*-maf*
DU	*-den*	*n-Ø*	*-Ø*	*-men*	*n- -m*	*-m*	*-maf*	*n- -maf*	*-maf*
PL	*-men*	*n--Ø*	*-Ø*	*-ben*	*n- -b/-m*	*-b/-im*	*-baf*	*n- -baf*	*-baf*

Table 2: Class 2 Argument suffixes in Marori

The Undergoer NP receives the U *=i* enclitic flagging, and is cross-referenced by the U prefix. The U prefix and the corresponding free pronouns are given in Table 3:

		1	2	3
Free Pronoun:	SG	*na*	*ka*	*efi*
	NSG	*nie*	*kie*	*emnde*
U Pref:		*i-* [2]	*k-*	*∅-*

Table 3: Free pronouns and S/O prefixes in Marori

2.2 Voice alternations

SUBJ and OBJ are relevant grammatical relations in Marori, distinct from semantic roles. Evidence for this comes from active-middle voice alternations in Marori. Consider the alternation in the following examples:

[2] In the intervocalic position *i-* becomes *ri-* (i.e. having an *r-* insertion), e.g. *ti=i-nggof* 'hide=1SG-AUX.NrPST' →*ti=ri-nggof.* In other phonological context, e.g. in the word-initial position it can appear as y-, and both i- and y- are equally acceptable, e.g. *yu-ngg-obon* / *i-ngg-obon* '1SG-AUX-1NPL.NrPST'.

(7) a. tege famndu mundwe=mi-du dwet=i.
fast very spin=3SG.M.U.AUX-1SG.A.PRES money=U
'I spin the coin very fast.' (ACTIVE)

b. tege famndu mundwa=n=ma-m dwet.
fast very spin=MID-AUX-3NPL.A.NrPST money
'The coin was spinning fast.' (MIDDLE)

Sentence (7a) represents the Active Transitive Voice, or Active for short, to be contrasted with Middle (7b) and Impersonal Active discussed below; cf. example (8). In the Active structure, the A argument is SUBJ and referential; this is realised by the verbal suffix *-du* in (7a). and referential. The patient argument (i.e., *dwet* 'coin') is OBJ. It receives the U 3.SG.M agreement on the verb (with the high vowel *i/e*) and it is flagged by the U clitic *=i*.

In the Middle structure in (7b), in contrast, there is no A argument present. The structure is morpho-syntactically intransitive. The NP *dwet* now shows up as the sole SUBJ argument, receiving the SUBJ suffix agreement *–m* and, crucially, no *=i* flagging.

Marori has no passive voice. The situation that requires a passive in other languages (e.g., an affected participant with an unknown agent) is expressed by either the (patientive) Middle or Impersonal Active construction. Consider this example of Impersonal Active construction:

(8) [Context: a boy who grabbed the horns of a deer, mistakenly thinking them as branches of a tree, and then was caught in between the horns.]
mbe sakud pende-fi tanduk rusa saune-fi.
exist hang 3SG.M.U.AUX-3A.RPST horn deer 3SG.M.put-3RPST
'...he got/was caught hanging, placed in between the deer horns.'

The understood patient argument in (8) is only expressed by the verbal U agreement morphology glossed as '3SG.M'. The verbal morphology is active in that it shows the third person A portmanteau suffix *-fi*. This suffix refers to no A participant in the context as seen from the translation. It is non-referential, having no argument linked to it at the level of argument structure. Its presence in the verbal morphology is primarily for TAM purposes. The analysis adopted in this paper for the case of Impersonal Active is that it is syntactically intransitive despite its active transitive morphology (with the third person A suffix). This is an instance of a mismatch between morphology and syntax, an issue further discussed in section 3.

2.3 Two types of middle constructions

The two types of middle constructions in Marori, MID-1 and MID-2, have their own properties, further elaborated in 2.3.1 and 2.3.2. They share the salient property, in contrast to the Active structure, of being morphologically intransitive, with SUBJ being the only argument indexed on the verb. Syntactically, however, a middle construction can be transitive: it comes with OBJ not indexed on the verb but possibly flagged with the U clitic *=i*. Middles in Marori also share cross-linguistically known properties of middles (Kemmer 1993); e.g. associated with inherently self-directed events (i.e. with an affected subject).

Which middle type a predicate can take is semantically determined in Marori. The broad pattern is this: verbs depicting inherently self-directed one-participant events such as *syepud* 'bathe' and *tpab* 'comb' and also psychological verbs such as *kamaen* 'angry' take MID-1; verbs whose affectedness is typically not self-oriented, canonically involving two-participants such as *tV-* 'hide' and *komow* 'wait', take MID-2.

2.3.1 Type 1 Middle (MID-1)

MID-1 in Marori is characterised by its AUX root form being *nggV*, and by its full constructed morphology, parasitic to the transitive U and A affixes. MID-1 can be schematised as follows:

(9) Type 1 NP X:PRED PREF:U_*i* – AUX:*nggV* – SUFF:A_*i*

Verbs in MID-1 include verbs that express self-indulgence, self-instigation, dynamic (psychological) states, reflexivity and reciprocity. Each of these is now exemplified.

Self-indulgence verbs include verbs such as *eni* 'play' and *ubun* 'camp'. The following are from a natural text in Marori, where the subject is realised by the coreferential U prefix and A suffix (in bold):

(10) sudah me kunonnjon, mbe sesei **yu**-nggo-**bon.**
already want go.home.PST exist work 1SG.U-AUX-1NPL.A.NrPST
'after I went home, I did some work.' (AWMarkus)

Verbs depicting self-instigation can be unergative (i.e. agentive), e.g. *syepud* 'dive, bathe' and *fedfed* 'squat', or possibly non-agentive, e.g. *kibib* 'roll'. They can be expressed in MID-1 structures:

(11) na fedfed yu-ngg-ra-du.
1SG squat 1SG.U-AUX-DUR-1SG.A.PRES
'I am squatting.'

Note that certain unergative verbs such as 'run' (e.g., *kundo-ru* 'run-1SG.FUT') are not expressed in middle morphology but in suffixing morphology. Thus there is some lexical idiosyncrasy in the classification of verbs into middles.

Dynamic (psychological) states such as *nggerngger* 'forget', *kamaen* 'angry' and *sira* 'afraid' belong to MID-1 too. Examples:

(12) a. na=i John=i nggerngger yu-ngg-obon.
1SG John=U forget 1SG.U-AUX-1NPL.A.NrPST
'I forget John.'

b. pa=na kamae =yu-nggo-ru.
soon-1SG angry=1SG.U-AUX.NPL -1SG.A.FUT
'I'll be angry.'

Verbs of grooming such as *tpab* 'comb' are in the middle. These verbs carry a reflexive or reciprocal meaning. The reciprocal meaning is imposed by the adverbial *endre-endre* 'in turn', without which the sentence would be ambiguous, with the other meaning being reflexive, 'we two combed our own hair'.

(13) na keke syepud i-ngg-ra-mon.
1SG there bath 1SG.U-AUX-DUR-1.A.NrPST
'I bathed (myself) (by diving) there.'

(14) nie endre-endre fa pu yar-nggwa-ra-den. cf. (1)
1NSG in.turn-REDUP with hair 1NSG.U-AUX-DUR-1DU.A.PRES
'We are two combing each other's hair.'

(15)

	Singular		Plural	
PERS:	Free pron	Refl.	Free pron	Refl
1	*na*	*namndu*	*nie*	*ninamndu*
2	*ka*	*kanamndu*	*kie*	*kinamndu*
3	*efi*	*ninamdu*	*emnde*	*anamndu*

Note that Marori has a distinct analytic reflexive construction making use of possessive reflexives shown in (15). An example is given in (16) which shows that the reflexive object is treated as a third person pronoun, as evidenced from the fact that it receives a third person object agreement in the form of zero prefix and low vowel /a/ glossed as '3SG'. (The first person U would have the prefix *i-* before the AUX *ma*, but is

unacceptable for the reflexive structure in (16).) As seen, the complex reflexive pronoun contains the noun head *mei* (literary meaning) 'body'.

(16) na namndu mei=i kaswa=ma-du.
1SG 1.REFL body=U 3SG.U.hit.PL=AUX-1SG.A.PRES
'I hit myself (Lit. I hit my own body).'

2.3.2 Type 2 Middle (MID-2)

Type 2 Middle whose construction is depicted in (17) has the following properties: i) the invariant prefix *n-* (instead of the inflected U prefix as in MID-1) in addition to the inflecting A suffix; ii) the AUX root varies depending on the lexical predicate (cf. the invariant root *nggV* for MID-1); and iii) the lexical predicate can be directly in MID-2 morphology. Each of these properties are exemplified below.

(17) Type 2: NP X:PRED PREF:*n_i* – AUX – SUFF:A_*i*

Consider the examples in (18) with the verb *komow* 'wait', which takes the auxiliary *mV*. Sentences (a) and (b) are intransitive in MID-2 structure with the invariant prefix *n-* and the inflecting A suffix (*-du*/*-den*) showing agreement with the subject. Sentence (c) is a transitive non-middle structure with the object being *ka* '2SG' receiving the U clitic *=i* and the verbal agreement *k-*.

(18) a. keke di=na komow n-ma-du.
here soon=1SG wait MID-AUX-1SG.A.PRES
'I just wait here.'

b. keke di=nie komow n-ma-den.
here soon=1NSG wait MID-AUX-1.DU.A.PRES
'We two just wait here.'

c. John nggowke ka=i komow k-ma.
John there 2SG=U wait 2SG.U-AUX.3NPL.A.PRES
'John waits for you there'

Sentences (19) exemplify the verb 'hide' which take the AUX *nggV*. Sentence (a) is intransitive middle with *n-* (i.e., MID-2) and reflexive meaning whereas the other two are active transitive with the verbs showing different U prefixes in agreement with the U NPs (*na* and *Tini*) marked by *=i*.

(19) a. na ti=n-nggo-bon. (MID-2 INTR.)
1SG hide-MID-AUX-1SG.A.NrPST
'I hid (myself).'

b. John na=i ti=ri-nggo-f. (ACTIVE TRANS.)
John 1SG=U hide=1SG.U-AUX-3.A.NrPST
'John hid me.'

c. Maria Tini=i ti=nggo-f. (ACTIVE TRANS.)
Maria Tini=U hide-3.U-AUX-3.A.NPL.NrPST
'Maria hid Tini.'

The verb *tfyV* 'meet, see' exemplifies the verb class that requires no inflected AUX and can appear in MID-2 structure. The verb appears in its transitive structure in (20a) with the masculine third person object Thomas. The verb agreement is reflected in the vowel *e*. The verb appears in intransitive MID-2 structure in (20b), with the subject being a coordinate NP with *fi*.

(20) a. na Thomas=i tefye-ben. (ACTIVE TRANS.)
1SG Thomas=U tfy.e-ben
meet.3SG.M-1NPL.A.NrPST
'I met Thomas.'

b. na Thomas fi tafanjabon. (MIDDLE INTR.)
1SG Thomas and tfy.a-n-bon
3.meet-MID-1NPL.A.NrPST
'Thomas and I met (with each other).'

3 LFG analysis

The parallel nature of the LFG model can nicely capture the intricacy of middle constructions in Marori, e.g. the intriguing case of middle in relation to the predicate *abon* 'steal', cf. example (40). The components of the proposed analysis consist of the nature of lexical entries, the specification of the verbal morphology and the morphology-syntax-semantic interface. Each will be outlined briefly below.

3.1 Lexical entries

Specifications in lexical entries in LFG make it possible to capture the following lexical constraints of Marori middle morphosyntax: a) the syntactic valence type, e.g. whether a predicate can appear in Type 1 or Type 2 Middle and b) the AUX selection by the lexical predicate, or the absence of it. Together with the c-str specification, we can also capture

the ultimate constructed transitivity of the predicate, e.g. morphologically middle-intransitive but syntactically transitive.

Consider the proposed sample entries for *ti* 'hide' and *eni* 'play' given in (21). They look the same in terms of argument structure: the patient argument is optional; i.e. allowing it to appear in transitive and intransitive syntax.[3] The selected AUX is also the same, namely *ngg*.

(21) a. *ti*
(↑PRED)= 'hide<A:agt, (P)>'
(↑AUX-FORM)=c *ngg*
(↑VAL-TYPE)= {MID-2 | ACTIVE}

b. *eni*
(↑PRED)= 'play<A:agt , (P)>'
(↑AUX-FORM)=c *ngg*
(↑VAL-TYPE)= MID-1

They significantly differ in the VAL-TYPE specifications, however. The predicate *ti* 'hide' allows a MID-2 and ACTIVE alternation (cf. example (19)) whereas *eni* 'play' is strictly of MID-TYPE 1. The predicate *eni* allows a transitive syntax but it must still have its verb in intransitive middle morphology, as shown in (22). This is different from *ti* 'hide' shown in (19b-c), where the transitive syntax also requires transitive morphology; i.e. the U prefix and A prefix must be referentially different.

(22) eni (bola=i) yu-ngg-ra-du.
play ball=U 1SG-AUX-DUR-1SG.PRES
'I am playing (football).'

The presence of the referentially distinct A and U triggers different valence structures and voice types for different verbs. For example, it requires the ACTIVE voice for the predicate *ti* 'hide' (cf. example (19b)) but it requires the MID-1 voice for *eni* 'play' as shown in (22).

Nevertheless, in both cases, the syntax is transitive: both A and P arguments in these two sentences are mapped onto SUBJ and OBJ respectively. Assuming a version of a-str and f-str mapping in LFG where linking is regulated by prominence matching of arguments (Arka 2003),[4] we can have the grammatical function mapping shown in (23) for the two sentences. The precise nature of the a-str of MID-1 reflecting the semantics of the middle voice will be further discussed below.

---

[3] In the interest of limited space, the valency in its its various forms (semantic, a-str and f-str) is given in a representation that does not fully comply with the standard LFG formalism, e.g. an optionality of P is shown as (P) rather than a disjunction.

[4] I assume a version of argument structure as described in Manning (1996) showing thematic prominence as well as syntactic prominence of coreness/obliqueness. The notations of A, R, and P/T reflect this information: A, R, and P/T are Actor, Recipient/Goal and Patient/Theme (macro-)roles respectively, ordered in that way.

(23) a. Active Transitive Construction, (19b):

SUBJ	OBJ
'ACTIVE.hide <A	P >'
'John'	'1SG'

b. Middle Transitive Construction, (22):

SUBJ	OBJ
'MID-1.play <A	P >'
'1SG'	'ball'

3.2 Verbal Morphology and function mapping.

I adopt a GPFM model (Stump 2001, Spencer 2010, 2013), where morphology with its M(orphological)-features is an autonomous module separate from syntax and semantics. The GPFM allows us to explicitly capture certain constraints of the complex interface between morphology-semantics-syntax, by means of different kinds of function mapping.

In this model, the transitive A and U agreement morphology carries M-features. For example, *i-* of *i*X has the paradigm function (PF) mapping shown in (24a), abbreviated in (24b). This says that, given the input with the stem/lexeme X and a set of M-features σ:{[ROLE:U], [PERS:1], [NUM:sg]}, the form *i*X is generated (with those relevant features) by the morphological system. Likewise, the forms X*du* and X*ru* are generated given the set of input features shown in (24c) and (24d) respectively. Note that there are additional TAM features {pres}/{fut} with these forms.

(24) a. PF(<X, σ:{[ROLE:U], [PERS:1], [NUM:sg]}>) = <***i***X , σ'>
b. PF(<X, σ:{U.1.sg}>) = <**i**X, σ'>

c. PF(<X, σ:{A.1.sg dur pres}>)= <X**du**, σ'>

d. PF(<X, σ:{A.1.sg fut}>)= <X**ru**, σ'>

The general morphology-syntax interface showing the morpho-syntactic constraints of ACTIVE and MID-1/MID-2 so far discussed can be captured as a PF mapping from morphology to syntax in a straightforward way, as shown in (25). The representation in (25a) says that the PF_{ACTIVE} has the input of A and U M-features with disjoint referential index, which are mapped onto a transitive valence structure. In contrast, the $PF_{MID\text{-}1}$ (b) has co-referential A and U and maps them to SUBJ. This captures the fact that the MID-1 verb is morphologically intransitive, though fully parasitic to the transitive morphology. $PF_{MID\text{-}2}$ is is different from like $PF_{MID\text{-}1}$ in having an invariant *n-*.

(25) VOICE TYPES AND MORPHOSYNTACTIC MAPPING IN MARORI:
i or *j* is an abbreviated index of a set of referential M-features {[PERS], [NUM], [GEND]}
y/*z* in *y*X*z* is a variant exponent depending on PERS, NUM & PERS
n in *n*X is an invariant *n*- exponent

	MORPHOLOGY:	SYNTAX:
a.	PF_{ACTIVE}(<X, σ:{A.*i* U.*j*} … >)	= < *y*X*z*, σ:{(↑SUBJ).*i* (↑OBJ)*j* …} >
b.	$PF_{MID\text{-}1}$(<X, σ:{A.*i* U.*i* …}>)	= < *y*X*z*, σ:{(↑SUBJ).*i* …}>
c.	$PF_{MID\text{-}2}$(<X, σ:{A.*i* U.*i*…}>) =	<*n*X, σ:{(↑SUBJ).*i* …}>

The notation *y* or *z* refers to the inflecting prefixal or suffixal exponent in yXz. The $PF_{MID\text{-}1}$ mapping in (26) illustrates the case where y and z are coreferential exponents associated with the root X=*nggV*. This shows the input-output process generating the form ***yu****nggo****bon*** in morphology. With the same input, the $PF_{GF.MID\text{-}1}$ mapping generates the associated set of functional equations shown in (27).

(26) $PF_{MID\text{-}1}$(<*nggV*, σ:{A.[1.sg]*i* U.*i* NrPST NonDur}>) = (*input*)
(<*yunggobon*, σ':{A.[1.sg]*i* U.*i* NrPST NonDur} (*output*)

(27) $PF_{GF.MID\text{-}1}$(<*nggV*, σ:{A.[1.sg]*i* U.*i* NrPST NonDur}>) = (*input*)
<*yunggobon,* σ':{(↑SUBJ)=*i* (↑TNS)=NrPST (↑ASP)=NonDur }> (*output)*

The next part of the analysis deals with the morphology-semantics interface to capture the marking and semantics of the middle. I propose that the middle M-features be mapped to the semantic structures reflecting at least three (cross-linguistically) common types of middle meanings, also reflected in the argument-structure. The three common meaning types of middles and their associated linking are shown in (28): (a) 'actions done for the joy/benefit of the doer', (b) 'self-affectedness/instigated action' (which covers reflexive and reciprocal meanings), and (c) P(atientive) orientation. I assume a model of semantic conceptual structure similar to that described in Jackendoff (1990).[5]

The notation [A=R]*i* and [A=P]*i* in (28) represents argument conflation (or 'argument unification') (cf. Rákosi 2008, Hurst 2010)[6]

[5] It remains to be worked out, however, how precisely the assumed conceptual structure is to be hooked up to the standard LFG architecture; see Butt (1995) for such a proposal, and also Schätzle and Butt (2015) for a linking schema for a middle based on Kibort (2013; 2014).

[6] The analysis here draws from work on reciprocals by Alisa (1996), Rákosi (2008) and Hurst (2010). However, instead of using Dowty's analysis of proto roles ([P-A][P-P]), I

where A, R and P are Actor-like arguments, Recipient-like arguments (including goal), and Patient-like arguments respectively. The index (*i*) shows argument identification and linking. For example, the most prominent argument in the a-str, [A=R] in (28a) is understood as 'affector' (i.e. the first argument of AFFECT), also the recipient (i.e. argument of FOR).

(28) a) SELF-INDULGENCE/BENEFIT
'PRED <[A=R]*i*, …>'
[AFFECT([*i*],[*j*]) FOR(*i*)]

b) SELF-AFFECTEDNESS
'PRED<[A=P]*i*>' (*a-str*)
[AFFECT([*i*],[*i*]) (*sem-str*)

c) P-ORIENTATION
'PRED<[P]*i* >' (*a-str*)
[AFFECT([ ],[*i*]) (*sem-str*)

Given (28), the PF that regulates the morphology-semantic interface, e.g. for the inflected middle form *yunggobon* (i.e. 'self-indulgence' middle), can be formulated in (29). For simplicity, only the output in the a-str is given here. The simultaneous effect of the operations of the PFs shown in (26), (27) and (29) is the generation of the middle AUX form *yunggobon* shown in (30) that comes with the f-str and a-str mapping constraint associated with the argument (index *i*) as well as other f-str specifications such as TNS and ASP.[7]

(29) $PF_{SEM.MID\text{-}1}$(<*nggV*, σ:{A.[1.sg]*i* U.*i* NrPST NonDur}>) = (*input*)
<*yunggobon,* σ':{[A=R]*i* }> (a-str *output)*

(30)	FORM:	F-STR:	A-STR:
	yunggobon	(↑SUBJ)= '1.sg'$_i$ (↑TNS)=NrPST (↑ASP)=NonDur	'PRED<[A=R]$_i$ …>'

When ***yu****nggo****bon*** combines with the predicate *sesei* 'work' as (10), whose argument structure is shown in (31), we get the f-str shown in (31b). The representation of 'work<(SUBJ:[A=R])>' in the f-str is a shorthand for the mapping of SUBJ and its [A=R] role, as the output of the

---

use generalised roles of A, R, and P commonly used by typologists (e.g., Haspelmath 2007) to capture a wider set of meanings of middles that include SELF-INDULGENCE/BENEFIT represented by [A=R]. The category of SELF-INDULGENCE/BENEFIT is roughly equivalent to what Kaufmann (2007) calls 'indirect reflexive'. However, I keep my category of self-indulgence to accommodate verbs such as *sesei* 'work' and *eni* 'play', which have no clear reflexive meaning.

[7] The representation such as '3.sg' abbreviates the standard equations in LFG such as (↑PERS) =3 and (↑NUM)=sg.

PF operations shown in (30). In short, we can capture the fact that, taken all together, the subtle meaning of *sesei yunggobon* (10) is actually 'I worked for the benefit of myself.'

(31) a. *sesei* (↑PRED) = 'work<A:agt>'
(↑AUX-FORM)=c *ngg*
(↑VAL-TYPE)= MID-1

b. *sesei yunggobon*

PRED	'work<(SUBJ)$_{[A=R]}$>'
SUBJ	'1.sg'
AUX-FORM	*ngg*
VAL-TYPE	MID-1
TNS	NrPST
ASP	NonDur

The analysis of structures that involve MID-2 can be done in the same way. The middle verb form to be generated is *n*X, and the semantics is typically of the self-affectedness/self-instigation type (largely with reflexive or reciprocal meaning) type (cf. (28b)). The PF rules for MID-2 can be formulated as in (32)-(34) below. Note that [A=P] in (34) is shorthand of the a-str/sem-str specifications shown in (28b).

(32) $PF_{MID\text{-}2}$(<X, σ:{A.*i* U.*i* …} >) = (*input*)
(<*n*X, σ:{A.*i* U.*i* …}> (*output*, morphology)

(33) $PF_{GF.MID\text{-}2}$(<X, σ:{A.*i* U.*i* …} >) = (*input*)
< *n*X, σ':{(↑SUBJ)=*i* … }> (*output*, syntax)

(34) $PF_{SEM.MID\text{-}2}$(<X, σ:{A.*i* U.*i* …} >) = (*input*)
< *n*X, σ':{[A=P]*i* }> (*output:* a-str/sem-str)

The rules correctly capture the salient property of MID-2: it results in an intransitive syntax because the A and P argument of the lexical predicate are conflated (34), and mapped onto SUBJ as in (33). For example, the predicate *pV* 'shave' must take *tnem* 'beard' as part of its predicate modification, and the P argument, *John*, is also understood as the possessor of *tnem*.

(35) a. John=i na tnem=pi-men. (ACTIVE TRANS.)
John=U 1SG beard=3SG.M.U.shave-1SG.A.DUR.NrPST
'I shaved John (Lit. I beard-shaved John).'

b. John tnem=nafram. (MID-2 INTRANS.)
John *tnem*=n-pa-ra-m
beard=MID-shave-DUR-3NPL.A.NrPST
'John was shaving (himself/his own beard).'

c. * John na=i tnem=naframm. (MID-2 TRANS.)
FOR: John was shaving me.' (cf. (35))

Two crucial properties should be noted from (35). Firstly, the agreement pattern provides evidence that the active-middle alternation (35a-b) involves a change in grammatical relation: the vowel of the predicate *pV* is inflected to become a high vowel /i/ showing the U/object M gender agreement in sentence (35). In the MID-2 form in sentence (b), in contrast, the vowel of the verb is /a/; i.e. a low vowel for non-3.SG.M argument if the argument is grammatically object. The fact that the masculine NP John can appear with the verb with /a/ means that this NP is not an object; in other words, sentence (b) is grammatically intransitive.

Secondly, the noun *tnem* in (35) is not a syntactic P/object. It is part of the predicate. Evidence for this comes from the fact shown in sentence (35a) where an overt object NP (*John)* is present. This object is referentially distinct from the subject; or else if there is no object as in sentence (35b) MID-2 must be used. Furthermore, MID-2 cannot have an overt object NP that is referentially distinct from the subject, as seen from the ungrammaticality of (35c).

The rules that generate MID-2 *naframm* with the associated properties are shown in (36)-(37). The lexemic root of *pV* 'shave' can be thought of as having the basic lexically-specified a-str information of (↑PRED) = 'shave<A P>'. It is also specified that it belongs to the MID-2 verb class.

(36) $PF_{GF.MID\text{-}2}$(<pV, σ:{A[3.NPL].*i* U.*i* } NrPst Dur Mid-2>) =
< *naframm,* σ':{(↑SUBJ)=[3.NPL].*i* (↑TNS)=NrPST
(↑ASP)=Dur (↑VAL-TYPE)= MID-2}>

(37) $PF_{SEM.MID\text{-}2}$(<pV, σ:{A[3.NPL].*i* U.*i* } NrPst Dur Mid-2>) =
< *naframm,* σ':{[A=P]*i* }>

Taken together, the PF rules (36)-(37) generate the form *naframm*, and a set of f-str/a-str information shown in (38). The system then correctly generates an intransitive structure where SUBJ is understood as both Agent and Patient, shown by the index *i*. The set of equations can then be part of the larger f-str information, e.g. of sentence (35b), shown in (39).

After outlining the proposed morphology-syntax-semantics interface of middle constructions in Marori, we are now ready to tackle what seems to be puzzling (at first) for the verb like *abon* 'steal' in Marori as seen in the following examples. Both sentences belong to MID-1, with the same middle form *yunggoru*. The syntax is different, however: intransitive (a)

vs. transitive (b). The A SUBJ *na* in both cases can also be flagged with the U clitic *=i*, even though it is not patientive.

(38)

FORM:	F-STR:	A-STR:
nafram	(↑PRED) 'shave<(↑SUBJ)*i* >	'shave<[A=P]*i* >'
	(↑SUBJ PERS)=3	
	(↑SUBJ NUM)=NPL	
	(↑TNS)=NrPST	
	(↑ASP)=Dur	
	(↑VAL-TYPE)= MID-2	

(39) F-str for sentence (35b):

$$\begin{bmatrix} \text{PRED} & \text{'beard.shave<(SUBJ)}_{\text{A=P}}\text{>'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'John'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{NPL} \end{bmatrix} \\ \text{TNS} & \text{NrPST} \\ \text{ASP} & \text{Dur} \\ \text{VAL-TYPE} & \text{MID-2} \end{bmatrix}$$

(40) a. na=i abon di=yu-nggo-ru. (INTRANS. MID-1)
1SG=U steal FUT=1SG.U-AUX-1SG.A.FUT
'I will steal (something).'

b. na=i bosik=i abon yu-nggo-ru. (TRANS. MID-1)
1SG=U pig=U steal 1SG.U-AUX-1SG.A.FUT
'I will steal pigs.'

The analysis in this paper is that *abon* 'steal' is of the 'self-benefit' type of middle predicate, whose basic argument structure is shown in (41). It says that *abon* 'steal' is a three-place predicate, as shown in an informal way in its sem-str in (41). Its a-str shows that *abon* is inherently middle of the self-benefit type, captured by the conflated [A=R] role specification (cf. (28a)). In addition, it is also specified that it belongs to the MID-1 class and that the thing stolen (P) is not required to be present as seen in (40); hence P is placed within brackets.

(41)

FORM:	F-STR:	A-STR:
abon	{(↑PRED)= 'steal<(↑SUBJ)$_i$ >	'steal<[A=R]$_i$ (P)$_j$>'
	|(↑PRED)= 'steal<(↑SUBJ)$_i$ (↑OBJ) > }	
	(↑VAL-TYPE) = MID-1	
A-STR:	'A_i does the stealing (of P) (for the benefit of R.self_i).'	

Given the principles of the morphology-syntax-semantics interface laid out earlier, the conflated [A=R] is mapped onto SUBJ and the auxiliary verb carries MID-1 marking *i-X-ru* realised as *yunggoru*, where both *i-* and *-ru* index SUBJ. Furthermore, the [A=R] conflation makes it possible for this argument to be flagged by *=i*, because it is actually R (i.e. an U property). Recall that a highly agentive argument without R or P conflation cannot not be flagged by the U clitic *=i* in Marori (cf. example (5)).

The proposed conflated [A=R] analysis of the a-str of the middle construction can account for cross-linguistic facts about the realisation of R in the middle. Since R is conflated to A, it is not realised as a distinct argument. It is understood as part of the meaning of the middle, e.g. as shown in (42b) from Fula (Niger-Congo) (Kaufmann 2007). In the proposed analysis it is also expected that a active-middle alternation may not change the grammatical relation involved, as in Ancient Greek: *loúô* X[acc] 'wash.1SG.ACT X = I wash X' vs. *loúômai* 'wash.1SG.MID= I wash X for myself' (Shibatani 2006:236).

(42) a. mi wu'y-ii=mo deptere. 'I lent him/her a book.'
I lend-PERF.ACT=3SG book (ACTIVE)

b. mi wu'y-ake deptere. 'I borrowed a book.'
I lend-PERF.MID book (MIDDLE)

Finally, the patientive middle (i.e. with the P-ORIENTATION Type, cf. (28c)) needs to be briefly discussed. Consider the pair of examples in (43). They show different voice types, with a subtle difference in meaning in terms of imperfectness of the DUR aspect, conceivable as the episodic/stage-level vs. individual-level distinction:

(43) a. na kara yu-ngg-ra-du. (MID-1)
1SG sick 1SG.U-AUX-DUR-1SG.A.PRES
'I am sick (at the moment of speaking).'

b. na kara yu-ngg-ra. (IMPERS. ACTIVE)
1SG sick 1SG.U-AUX-DUR.3NPL.A.PRES
'I am (often) sick (not necessarily at the moment of speaking).'
(or, 'I am often affected by sickness.')

c. PRED 'sick<[P]*i* >' (*a-str*)
[AFFECT([ ],[*i*]) (*sem-str*)

With the MID-1 structure (43a), it refers to progressive or on-going dynamic process, whereas in the Impersonal Active one (43b) it refers to

the habitual dynamic process. As far as the a-str is concerned, in both cases the subject *na* '1SG' is patientive, having no control over the sickness. They both share the same a-str/sem-str shown in (43c).

Note that the suffix –du carries the M-feature of {A} at the level of morphology. However, this feature has no contribution in terms of agentivity at the level of sem-str, and correspond to no A argument at the level of a-str. The only relevant contribution of the features of *-du* is in relation to ASP and TNS.

However, this is not always the case. With other predicates that depict events with a certain degree of controllability, e.g. 'cough', the alternation between MID-1 and Impersonal Active does give rise to a subtle difference in agentivity. Consider the pair of examples below:

(44) a. medi mar kafra-ru, pa toufo=ri-nggo.
if.FUT NEG drink-1SG.FUT FUT cough=1SG.U-AUX.3PRES.
'If I don't drink, I will cough.' (i.e. unintentional only)

b. na taufa=ri-ngg-ra-du.
1SG cough=1SG.U-AUX-DUR-1SG.A.PRES
'I am (deliberately) coughing now.' (possibly intentional)

Sentence (44a) is in Impersonal Active (with 'coughing' understood as unintentional only) whereas sentence (44b)is in MID-1 (with 'coughing' understood as possibly intentional).

I propose that certain predicates are inherently classified as highly patientive, while others such as 'cough' are canonically patientive but they allow certain degree of agentivity; hence partly controllable, depicted by a conflated [(A=)P] role. Thus, 'cough' in Marori is essentially intransitive (a one-place predicate) whose lexical entry is shown in (45). This lexical specification is consistent with the MID-1 meaning of 'self-affectedness' shown in (28b), and the deliberate interpretation of 'coughing' is constructionally imposed when the predicate combines with MID-1 AUX as seen in (44b). This is literally interpreted as 'I coughed myself'. In addition, since the A conflation is optional (indicated by [(A=)P]) we can also capture that, even when in MID-1, the coughing event can still be understood as totally patientive (cf. the free translation in (44b)). In this case, the P-orientation meaning of MID-1 applies.

(45) toufow
PRED 'cough<[(A=)P]*i* >' (*a-str*)
[AFFECT([ ],[*i*]) (*sem-str*)

The non-middle structure (44a) can be also straightforwardly captured. When *toufow* combines with the AUX in Impersonal Active structure, while the A M-feature (glossed as '3PRES') is arguable present in the morphology, given the model adopted here, it corresponds to no A argument in the semantic structure. Note that there is no argument in the first argument position of AFFECT in (45).

To conclude, possible subtle differences in meaning in the middle and Impersonal Active alternation is a result of interplay between the lexical information of a predicate and the information constructionally imposed the verbal morphology.

4 Final remarks

This paper has discussed middle constructions in Marori. Two salient properties of middles in this language include the following properties. Firstly, there are two middle constructions. Middle Type 1 (MID-1) is fully parasitic to transitive A/U inflectional morphology whereas Middle Type 1 (MID-2) is only partially parasitic, making use of the invariant *n-* instead of the U prefix. The distribution of the two is largely lexically determined, even though there is evidence that MID-2 is derivational in nature (e.g., turning a transitive stem to intransitive, the non-finite MID-2 verb retaining its *n-* prefix). Secondly, middles in Marori provide good evidence for the idea that ultimate clausal structures are morpho-syntactically constructed. Morphologically, this is clear from the morphological make-up of MID-1 verb, illustrated by example (4): intransitive morphology is constructed by having co-referential A and U exponents on the same verb. However, the morphologically intransitive verb can enter a transitive construction with the presence of the object NP analytically flagged by *=i*.

These two empirical points on Marori middle constructions pose a theoretical challenge in the analysis of middles: lexical (e.g., Fagan 1992), syntactic (e.g., Hoekstra and Roberts 1993), constructional (e.g., Iwata 1999), or a parametric combination of these (Marelj 2004). In this paper, I have demonstrated that LFG is well equipped to handle Marori middles, and that an LFG-based lexical-constructional analysis has been proposed. The analysis makes use the full machinery of LFG rich lexical specifications and parallel-based model, augmented by the mechanism from GPFM. It has been demonstrated that the morphology-syntax-semantics interface involved in the complex expressions in Marori can be captured explicitly within GPFM, in particular the idea of morphology as an autonomous module in grammar, separate from syntax and semantics.

5 References

Alsina, Alex. 1996. *The Role of Argument Structure in Grammar: Evidence from Romance.* Stanford: CSLI Publications.

Arka, I Wayan. 2003. *Balinese morphosyntax: a lexical-functional approach*. Canberra: Pacific Linguistics.

Butt, Miriam. 1995. *The Structure of Complex Predicates in Urdu.* Stanford: CSLI Publications.

Dalrymple, Mary, M Kanazawa, Y Kim, S Mchombo, and S Peters. 1998. "Reciprocal Expressions and the Concept of Reciprocity." *Linguistics and Philosophy,* 21 (2):159-210.

Fagan, S. 1992. *The Syntax and Semantics of Middle Constructions: A Study with Special Reference to German*. Cambridge Cambridge University Press.

Haspelmath, Martin. 2007. "Ditransitive alignment splits and inverse alignment." *Functions of Language,* 14 (1):79–102.

Hoekstra, Teun, and I Roberts. 1993. "Middle constructions in Dutch and English." In *Knowledge and language.* edited by E. Reuland and W. Abraham, 183–220. Dordrecht: Kluwer.

Hurst, Peter. 2010. "The syntax of lexical reciprocal constructions." In *Proceedings of LFG 2010*, edited by Miriam Butt and T. H. King. CSLI Publications. http://csli-publications.stanford.edu/

Iwata, Seizi. 1999. "On the status of implicit arguments in middles." *Journal of Linguistics*, 35:527-553.

Klaiman, M. H. (1991). *Grammatical voice*. Cambridge: Cambridge University Press.

Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge, Mass.: MIT Press.

Kaufmann, Ingrid. 2007. "Middle voice." *Lingua*, 117:1677-1714.

Kemmer, Suzanne. 1993. *The middle voice*. Amsterdam: John Benjamins.

Manning, Christopher D. 1996. *Ergativity: argument structure and grammatical relations*. Stanford: CSLI Publications.

Marelj, Marijana. 2004. *Middles and Argument Structure across Languages. Doctoral Dissertation*, Utrecht Institute of Linguistics OTS.

Rákosi, György. 2008. "The inherently reflexive and the inherently reciprocal predicate in Hungarian: Each to their own argument structure." In *Reciprocals and reflexives: theoretical and typological explorations*, edited by E. König and V. Gast. Berlin, Germany: Mouton de Gruyter.

Schätzle, Christin, & Butt, Miriam. 2015. "The diachrony of dative subjects and the middle in Icelandic: a corpus study.*"* in the Proceedings of the LFG15 Conference, edited by Miriam Butt and T. H. King. CSLI Publications, http://csli-publications.stanford.edu/

Spencer, Andrew. 2013. *Lexical relatedness: a paradigm-based model*. Oxford: Oxford University Press.

Stump, Gregory T. 2001. *Inflectional Morphology: a theory of paradigm structure*. Cambridge: Cambridge University Press.

NUMBER MISMATCHES IN COORDINATION AN LFG ANALYSIS

Oleg Belyaev
Lomonosov Moscow State University

Mary Dalrymple
University of Oxford

John J. Lowe
University of Oxford

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King (Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

In some languages, including Russian and Italian, it is possible for a plural noun to be modified by two or more coordinated singular adjectives. At the same time, it is possible for a singular noun to be modified by two or more coordinated singular adjectives, but for the reference of the noun phrase to be plural, rather than singular. We provide a formal analysis of these problematic agreement phenomena, making use of the distinction between INDEX and CONCORD agreement features, and the distinction between distributive and non-distributive features. We propose that the distributivity or non-distributivity of a feature may vary not only between different languages, but even on a construction-by-construction basis within a language. Specifically, CONCORD is a non-distributive feature in certain constructions in languages like Russian and Italian.

1 Introduction

Agreement is usually understood as the covariance of a formal property of one element with a semantic or formal property of another element (Steele 1978: 610; Corbett 2006: 4). On this view, standard patterns of agreement between attributive adjective and modified noun, i.e. where adjective and noun show the same features, are very simply dealt with. However, this is challenged by certain patterns in adjective coordination. In Russian and Italian, among other languages, it is possible for a plural noun to be modified by two or more coordinated singular adjectives. In such cases, the conventional approach to agreement does not seem to work: not only do the features on the controller and target differ, but it is the "target" (the adjectives) that seems to determine the number features of the "controller" (the noun), and not vice versa. This possibility exists alongside the possibility of using one or more coordinated plural adjectives to modify a plural noun. In other languages, such as Hindi, number agreement between adjective and head is obligatory, so it is not possible to use coordinated singular adjectives to modify a plural noun.

Although such patterns have been described in the literature, no explicit theoretical analysis has been proposed which can capture either the language-specific patterns, or the broader typological variation. In this paper, we will show how the approach to agreement adopted in LFG can be modified to handle all of the problematic cases. Our analysis makes use of the distinction between CONCORD and INDEX agreement features (Wechsler and Zlatić 2003), and the distinction between distributive and non-distributive features (Dalrymple and Kaplan 2000). Typically,

[†]We are very grateful to Miriam Butt and Farhat Jabeen for the Hindi data, to Martin Maiden, Alessandro Carlucci, Silvio Cruschina, Stella Merlin Defanti, Norma Schifano, and Sandra Paoli for their help with the Italian data, and to Bozhil Hristov for helpful discussion. We thank the audiences at the 4th conference on Constructional & Lexical Semantic Approaches to Russian, St Petersburg, and at LFG15, Tokyo, for the attention and comments, in particular Maria Kholodilova, Ron Kaplan, Agnieszka Patejuk, and Miriam Butt. John Lowe acknowledges the financial support of the Leverhulme Trust. Oleg Belyaev acknowledges the support of the Russian Science Foundation, project no. 14-18-03270.

CONCORD is analyzed as a distributive feature while INDEX is treated as a non-distributive feature (King and Dalrymple 2004), but we propose that CONCORD may, in certain cases, be non-distributive. We propose that the distributivity or non-distributivity of a feature is not only subject to variation between different languages, but even between different syntactic constructions/patterns within a single language. Our proposal has been implemented in XLE and demonstrated to work for all the data under discussion here.

In the following section, we discuss the data from the languages in question. In §3, we introduce the theoretical and formal assumptions underlying our analyses, which we present in §4. In §5 we discuss further issues, and in §6 we draw our conclusions.

2 The data

2.1 Italian

In Italian, attributive adjectives generally show agreement in number and gender with the noun they modify. So in (1), two coordinated singular adjectives agree with the singular head noun, while in (2) two coordinated plural adjectives agree with the plural head noun.

(1) *la* *vecchia* *e* *piccola* *stazione*
the.SG old.SG and small.SG station.SG
'the old and small station' (Italian)

(2) *le* *vecchie* *e* *piccole* *stazioni*
the.PL old.PL and small.PL station.PL
'the old and small stations' (Italian)

On its most prominent reading, example (1) involves a 'joint' reading of coordinated modification (Heycock and Zamparelli 2005): there is one single station, described as both old and small.[1] A joint reading is also available in (2), which may be understood as denoting several stations, each of which is both old and small. Under the joint reading of coordinated modification, it is not possible for two or more singular coordinated adjectives to modify a plural noun (3), and under no circumstances is it possible for two or more coordinated plural adjectives to modify a singular noun (4).

(3) **la/*le* *vecchia* *e* *piccola* *stazioni*
the.SG/the.PL old.SG and small.SG station.PL
Intended: 'the old and small stations' [each station is both old and small]
(Italian)

[1] We note the existence of another reading, the 'split' reading, in Section 2.1.2: (1) can also refer to two stations, one old and one small.

(4) **le/*la* *vecchie* *e* *piccole* *stazione*
the.PL/the.SG old.PL and small.PL station.SG
Intended: 'the old and small station(s)' (Italian)

2.1.1 Resolving agreement

However, there is another reading available for coordinated modification: the 'split' reading (Heycock and Zamparelli 2005). So, example (2) is in fact ambiguous: the reference may be to a single set of stations which are both old and small (the joint reading), or to two separate sets of stations, one set old, the other set small (the split reading).

Under this reading, it is possible for a plural noun to be modified by two or more coordinated singular postnominal adjectives (5). In (5), the phrase *le bandiere rossa e bianca* refers to two flags, with the attributes 'red' and 'white' each holding of a different flag. There is no direct number agreement between each adjective and the noun, but on a more abstract level there is a correlation between the total number indicated by the coordinate adjective set and the number marking on the noun.[2]

(5) *Alla* *partenza* *saranno* *ammainate* ***le*** ***bandiere*** ***rossa*** ***e***
to departure will.be.PL lowered.PL the.PL flag.PL red.SG and
bianca *accompagnate* *possibilmente* *da* *segnale* *acustico.*
white.SG accompanied.PL possibly by signal acoustic
'At the departure **the red and white flags** will be lowered, possibly accompanied by an acoustic signal.' [2 flags total: one red, one white] (Italian[3])

In semantic terms, this pattern is notable: the number marking on the adjectives makes a very clear semantic contribution to the interpretation of the phrase. It is only the number marking on each adjective that determines the absolute cardinality of each conjunct, and thus determines the cardinality of the whole noun phrase. We refer to this pattern as **resolving** agreement.

2.1.2 Non-resolving agreement

In Italian, this is not the only agreement possibility. With prenominal adjectives, the resolving pattern is unavailable; instead, all coordinate adjectives, as well as the head noun, have the same number features. The function of the number marking on the adjectives remains the same, but the number marking on the noun does not

[2]There is no dual in Italian, i.e. PLURAL indicates any number greater than 1. In Sanskrit, a language which has a dual and licenses this pattern of agreement, two singular coordinated adjectives modify a noun in the dual, not the plural, and we expect that similar patterns should hold also in other languages with more complex number systems and resolving agreement.

[3]`http://www.primazona.org/Avviso_VeleggiataSanGiovanni_2015.htm`

reflect the cardinality of the whole noun phrase. That the noun phrase as a whole has a plural number feature is clear from the plural verb agreement.

(6) *Prima del 1991, un giocatore non poteva giocare per la sua*
before of a player not could play for the his
nuova squadra finché ***la vecchia e nuova società*** *non*
new team until the.SG old.SG and new.SG club.SG not
si fossero accordate sulla cifra del trasferimento.
REFL were.PL agreed.PL on.the number of transfer
'Before 1991, a player could not start playing for a new team before **the old and the new club** had agreed on the amount (paid) for the transfer.' [2 teams total: one old, one new] (Italian[4])

(7) *La novità era nel senso che essa cambiava la natura*
the novelty was in.the sense that it changed the nature
della liquidazione, cosicché ***vecchio e nuovo regime***
of.the liquidation so.that old.SG and new.SG regime.SG
diventavano non più comparabili ...
became.PL not anymore comparable
'The novelty was in the sense that it changed the nature of liquidation, so that the **old and new regimes** became no longer comparable ...' [2 regimes: one old, one new] (Italian, La Repubblica corpus)

We refer to this pattern as **non-resolving** agreement.

As long as the adjectives concerned are not semantically incompatible, it is of course possible that all-singular agreement can give rise to a joint reading referring to a single individual, in which case the number of the noun phrase is determined by the head noun, and the number marking on each adjective is not counted cumulatively in determining the number of the whole. In the examples of all singular agreement given above, (6) and (7), the adjectives involved are semantically incompatible, so such a reading is impossible. Both readings are in fact available in (1), which may refer either to a single station that is both old and small, or to two different stations, one of which is old and one of which is small. The noun phrase-internal agreement pattern seen in these all singular examples is essentially the same as in the all plural example in (2) which, like (1), has both joint and split readings; as shown in examples (6) and (7), the verb shows plural agreement in the split reading, while in the joint reading the verb shows singular agreement.

It should be noted that mixed singular and plural conjuncts in the non-resolving pattern are not allowed, thus the description of this type as involving agreement between all elements is valid:

(8) **la vecchia e nuove biblioteca*
the.SG old.SG and new.PL library.SG
Intended: 'one old and several new libraries' (Italian)

[4] http://europa.eu/rapid/press-release_IP-95-1411_it.htm

2.1.3 Summary

A summary of the agreement types that are available in Italian is provided in the following table:

(9)

position	A1	A2	N	split	joint
prenominal	SG	SG	SG	+	+
	PL	PL	PL	+	+
	(SG	SG	PL	−	−)
postnominal	SG	SG	SG	−	+
	PL	PL	PL	+	+
	SG	SG	PL	+	−

The generalization regarding adjective position is fairly robust for Italian. Among 1,000 examples of "ADJ *e* ('and') ADJ NOUN" sequences randomly selected from the La Repubblica corpus (Baroni et al. 2004), we have found 18 clear cases of non-resolving agreement and no cases of resolving agreement. Inversely, for 1,000 random examples of "NOUN ADJ *e* ('and') ADJ" sequences, there are 13 cases of resolving agreement and no cases of non-resolving agreement. Therefore, while individual variation is possible (as suggested by reviewers), the distribution in question seems to be a feature of core Italian grammar.

2.2 Russian

Italian is not the only language that displays the two patterns. Russian largely shows the same agreement rules as Italian:

(10) *vysokij i xudoj mužčina*
tall.SG and thin.SG man.SG
'(A/the) tall and thin man' [joint reading, 1 man total]
'(A/the) tall man and (a/the) thin man' [split reading, 2 men total] (Russian)

(11) *vysokie i xudye mužčiny*
tall.PL and thin.PL man.PL
'(The) tall and thin men' [joint/split reading, 2 or more men total] (Russian)

(12) *vysokij i xudoj mužčiny*
tall.SG and thin.SG man.PL
*'(The) tall and thin men' [joint reading]
OK: '(The) tall and thin men' [split reading, 2 men total] (Russian)

(13) **vysokie i xudye mužčina*
tall.PL and thin.PL man.SG
Intended: '(The) tall and thin man/men' (Russian)

Like Italian, Russian displays both the resolving and non-resolving agreement types in adjective coordination with a split reading. But since all adjectives in Russian are prenominal (except for a few special cases), the patterns are (in general) freely interchangeable, not being structurally restricted as is the case in Italian.

2.2.1 Resolving agreement

Resolving agreement in Russian is illustrated by the split reading of (12), and the following noun phrase:

(14) *krasnyj* *i* *belyj* *flagi*
red.SG and white.SG flag.PL
'(the) red and (the) white flag' [2 flags total: one red, one white] (Russian)

Note that, as in Italian, the absolute cardinality of the noun phrase is determined on the basis of the cardinality of the adjectives: each singular adjective can refer to only one flag, so the cardinality of the phrase must be 2. In the following example too, it is the number marking on each adjective that indicates the cardinality (2+) of each conjunct, and it is only on the basis of that that the 4+ cardinality of the whole phrase can be inferred (since the plural marking on the noun indicates only a total cardinality of 2+).[5]

(15) *krasnye* *i* *belye* *flagi*
red.PL and white.PL flag.PL
'(the) red and (the) white flags' [4+ flags total: 2+ red, 2+ white] (Russian)

The resolving agreement type involves a kind of natural coordination effect in Russian (Wälchli 2005, Dalrymple and Nikolaeva 2006):

(16) *dobryj* *i* *zloj* *policejskie*
good.SG and evil.SG policeman.PL
'good cop and bad cop' [referring to an interrogation technique] (Russian)

(17) *??dobryj* *i* *zloj* *sosedi*
good.SG and evil.SG neighbour.PL
intended: 'good neighbour and evil neighbour' (Russian)

2.2.2 Non-resolving agreement

Non-resolving agreement in Russian is generally freely interchangeable with the resolving pattern. The following examples illustrate the non-resolving pattern:

[5]The split reading is the only possibility for the plural equivalent of example (15), because coordinated colour adjectives cannot have a joint reading in Russian. For the joint reading, a compound modifier would be used, e.g. *krasno-bel-yj* (red-white-M.SG.NOM) 'red and white'.

(18) *staryj i novyj stil' budut uravneny*
old.SG and new.SG style.SG become.FUT.PL equal.PL
'The old and new styles will become equal.' [2 styles: one old, one new] (Russian, Russian National Corpus (RNC))

(19) *staryj i novyj obraz stali nakladyvat'sja*
old.SG and new.SG image.SG begin.PST.PL superimpose.INF
'The old and the new image began to superimpose themselves.' [2 images: one old, one new] (Russian, RNC)

If anything, non-resolving agreement is the least marked construction of the two, because the natural coordination effect observed for the resolving pattern does not seem to hold for non-resolving agreement:

(20) *u menja na dače byli dobryj i zloj sosed*
at me in dacha were.PL good.SG and evil.SG neighbour.SG
'At the dacha I had a good neighour and a bad neighbour.' (Russian)

2.2.3 Summary

Since Russian does not distinguish between adjective positions, the summary is simpler than in Italian, but in essence is identical:

(21)

A1	A2	N	split	joint
SG	SG	SG	+	+
PL	PL	PL	+	+
SG	SG	PL	+	−

2.3 Hindi

Agreement mismatches such as these are not possible in all languages. In Hindi, coordinated adjectives agree in number with the head noun; so coordinated adjectives modifying a plural noun must appear in the plural, even if each adjective refers to a set of cardinality one (22). The only way to get the 'exactly one of each' reading is to use a singular noun, with coordinated adjectives in agreement (23).

(22) *ye hare aur piile jhaṇḍe*
this.PL green.PL and yellow.PL flag.PL
'these green and yellow flags' [split reading: 1+ flags of each colour]
'these green and yellow flags' [joint reading: 2+ part green, part yellow, flags] (Hindi)

(23) *yah* *haraa* *aur* *yah* *piilaa* *jhaṇḍaa*
this.SG green.SG and this.SG yellow.SG flag.SG
'this green and this yellow flag' [2 flags total] (Hindi)

The summary for Hindi is thus such that it allows the split reading only in all-plural contexts:

(24)

A1	A2	N	split	joint
SG	SG	SG	−	+
PL	PL	PL	+	+
(SG	SG	PL	−	−)

2.4 Previous analyses

Although such patterns have been described in the literature, no explicit theoretical analysis has ever been proposed to capture either the language-specific patterns or the broader typological variation. The Russian data have been described and discussed in Kodzasov (1987) and Iomdin (1990) in the framework of Meaning-Text Theory. Kodzasov provides an extensive description of the relevant patterns and the semantic restrictions that they involve, but gives only a sketch of a possible syntactic analysis. Iomdin (1990) reviews several potential syntactic analyses in terms of dependency grammar. He concludes that none is satisfactory, and the attested patterns should be relegated to semantics instead (an analysis of which is not explicitly described).

It is noteworthy that Iomdin describes both structures as involving what he terms *sočinitel'noe sokraščenie* ("coordinating reduction") which, in the Russian tradition, is a term roughly analogous to ellipsis. It is certainly appealing to view examples like (18)–(19) (or 6–7 in Italian) as involving ellipsis, i.e. 'the old ~~station~~ and new station'. However, an ellipsis account is not satisfactory for (14) and (16) (or 5 in Italian): the noun is not recoverable, due to the mismatching number. Therefore, an ellipsis account would fail to explain all the data. It is also difficult to see why the availability of ellipsis would depend on the adjectives' position in Italian.

Furthermore, noun agreement seems to be the only property differentiating ellipsis from non-ellipsis on this view. Other properties do not vary with the number of the noun. For example, both constructions allow non-constituents to be coordinated:

(25) *[[graždanskogo* *atomnogo]* *i* *[voennogo*
civil.GEN.SG nuclear.GEN.SG and military.GEN.SG
raketno-jadernogo] *flota]*
rocket-nuclear.GEN.SG fleet.GEN.SG
'of the civil(sg.) nuclear(sg.) and the military(sg.) nuclear-rocket(sg.) fleet(sg.).' (Russian, RNC)

(26) *v [[Moskovskom sel'skoxozjajstvennom] i*
in Moscow.ADJ.PREP.SG agricultural.PREP.SG and
[Kievskom politexničeskom] institutax]
Kiev.ADJ.PREP.SG polytechnic.PREP.SG institutes.PREP.PL
'at the Moscow Agricultural and the Kiev Polytechnic Institutes' (Russian, RNC)

Thus, a satisfactory analysis must treat the two constructions together.[6]

Bosque (2006) takes a different approach. He discusses Spanish examples like *los embajadores mexicano y argentino* (the.PL ambassadors.PL Mexican.SG and Argentinian.SG), which are analogous to (5) in Italian. His solution is to assume that the coordinate adjectives are in fact DPs with null pronominal heads, i.e.: the.PL ambassadors.PL [*pro* Mexican.SG] and [*pro* Argentinian.SG]. The noun then agrees with the resolved plural feature of the coordinate DP. As we will see, the core idea of this analysis is similar to our approach. But it does not explain why non-resolving examples such as (6) or (18) are possible: if each adjective heads a DP, there should be no singular agreement.

Below, we will show how the data can be accounted for without resorting to null pronouns or relegating the matter to semantics.

3 Agreement

Agreement is sometimes defined as variation in the value of a formal property of one element, e.g. case, number or gender, depending on the value of that property in another element. In other words, the 'target' of agreement merely reflects the formal properties of the 'controller' of the agreement, and does not make a semantic contribution of its own. For example, in the Russian phrase

(27) *ja tebja vižu*
I you.SG.ACC see.1SG.PRES
'I see you.'

the 1st person and SG features of the verb (the 'target') can be analysed as determined by the inherent 1st person and SG features of the subject pronoun (the 'controller'). This approach can easily deal with standard patterns of agreement, as in the phrase immediately above, but it is more difficult to see how it could deal with the patterns discussed in the previous section, where the number of the adjectives does not reflect the number of the noun head. Indeed, in such examples the number marking on the adjectives has the role of specifying the number of each conjunct. In the case of 'non-resolving' agreement between coordinated singular

[6] A full analysis of non-constituent coordination is beyond the scope of this paper. In the context of LFG, the finite state approach of Maxwell and Manning (1996) can probably provide a solution.

adjectives with a split reading and a singular noun, it would be possible to treat the singular number marking on the adjectives as directly determined by the singular noun, but it then remains to explain why the singular number of the noun does not reflect the plurality of the noun phrase as a whole, as shown by verb agreement.

The standard approach to agreement within LFG (as also in HPSG, and some typological literature, e.g. Kibrik 2011) is to assume a codescriptive approach. Under this approach, the 'target' and 'controller' both contribute to determining the number of the controller. For example, the lexical entry for the Russian pronoun *ja* 'I' contains the specifications in (28), while the lexical entry for *vižu* 'see' includes those in (29).

(28) *ja* (↑ PRED) = 'I'
(↑ PERS) = 1
(↑ NUM) = SG

(29) *vižu* (↑ PRED) = 'see'
(↑ SUBJ PERS) = 1
(↑ SUBJ NUM) = SG

When the pronoun *ja* 'I' functions as subject to the verb *vižu* 'see', both verb and pronoun specify the features of the subject's f-structure. As we will show, this codescriptive approach is able to deal with the complicated agreement patterns discussed in the previous section.

Our analysis also relies on the distinction between INDEX and CONCORD features as two distinct types of agreement features. Wechsler and Zlatić (2003), working within HPSG, show that nouns distinguish two types of agreement features: CONCORD features, which generally control agreement within a noun phrase, e.g. between a noun and any determiners or modifying adjectives; and INDEX features, which generally control noun phrase external agreement, e.g. between a noun phrase and an agreeing verb. Building on this work, King and Dalrymple (2004) explore the INDEX/CONCORD distinction in agreement with coordinated singular nouns:

(30) [This/*these man and woman] are/*is eating sushi.

The coordinate noun phrase in (30) consists of two singular nouns. The determiner is required to appear in the singular, but the verb is constrained to appear in the plural, suggesting that the CONCORD (noun-phrase internal) and INDEX (noun-phrase external) agreement features of the noun phrase have different values.

Finally, our analysis is based on the distinction between distributive and non-distributive features (Dalrymple and Kaplan 2000). Some f-structure features are non-distributive, which means that the feature is associated with the set representing a coordinate structure independent of the features of the individual members of that set. Other f-structure features are distributive, which means that a feature specified for a set representing a coordinate structure can only be associated with the individual members of the set, and not with the set itself; crucially, the value of the feature must be the same for all conjuncts. Dalrymple and Kaplan (2000) show that any requirement made of a set in relation to a distributive feature is satisfied if the requirement holds of every member of that set.

4 Proposal

As mentioned above, CONCORD agreement features are typically analysed as distributive, and as having relevance for noun phrase internal agreement, while INDEX agreement features are typically treated as non-distributive, and as having relevance for noun-phrase external agreement. King and Dalrymple (2004) argue that in some languages, including Russian, INDEX may also have relevance for noun-phrase internal agreement, due to agreement patterns such as the following:

(31) *moi* *brat* *i* *sestra* *prišli*
my.PL brother.SG and sister.SG come.PST.PL
'My brother and sister came.' (Russian)

In this example, both nouns are singular, so must have singular CONCORD, yet the determiner shows plural number. The determiner therefore appears to depend on the plural INDEX feature of the coordinate set, rather than the CONCORD features of the conjuncts. However, this will not suffice to explain the patterns seen in (18)–(20): the plural verb marking shows that the INDEX feature of the coordinate set is plural, yet all elements of the noun phrase are singular. Example (18) is repeated here as (32).

(32) *staryj* *i* *novyj* *stil'* *budut* *uravneny*
old.SG and new.SG style.SG will.be.PL equal.PL
'The old and new styles will become equal.' [2 styles total] (Russian)

We propose that the agreement patterns discussed above can be explained without altering the original assumptions regarding the distinction between INDEX and CONCORD: noun-phrase internal agreement is determined purely by reference to CONCORD, while noun-phrase external agreement is determined purely by reference to INDEX. However, we do require an alternative approach to the distributivity of these features. Specifically, we propose that CONCORD can, in certain circumstances, be non-distributive. While INDEX is universally non-distributive, CONCORD is subject to variation across languages, and even across different constructions within particular languages.

Specifically in relation to the Italian and Russian data, the contrasts of (5) vs. (6) and (14) vs. (18) can be explained by assuming that CONCORD is non-distributive (and resolved as plural) in (5) and (14) but is distributive (and thus forced to be the same for all conjuncts) in (6) and (18). Thus, the distributivity of CONCORD varies in Russian and Italian on a construction-specific basis (in Italian this is linked to word order patterns, but word order is not relevant to the Russian data). For languages like Hindi (22–23) and English, on the other hand, CONCORD features are purely distributive.

It is thus possible to treat typological variation in the status of CONCORD by reference to a 'cline of distributivity', with CONCORD always distributive at one

extreme, and movement along the cline correlated with increasing nondistributivity on a construction-by-construction basis:[7]

(33) Distributivity of CONCORD:

always d.		sometimes d.		often non-d.
English/Hindi	>	Italian	>	Russian

The distributivity or otherwise of a feature is not in principle variable in LFG. There is no way to specify or change the distributivity of a feature in either the lexicon or the syntax (or anywhere else): it is an inherent, pre-established property of a feature.[8] In proposing that CONCORD can be non-distributive in only some constructions in a language, we therefore require some way to simulate this variation without actually changing the status of the feature. We do this by assuming that if in a language CONCORD *can* be non-distributive, then this is its pre-specified status; the effects of distributivity are enforced, in those constructions where CONCORD appears to be distributive, by annotations on phrase-structure rules which require the CONCORD value of the coordinate structure to be the same as the CONCORD value of each conjunct.

We assume that languages may have different coordination rules for adjective phrases, one the normal set-forming coordination rule (giving the joint reading), and the other creating separate f-structures, with each adjective appearing as a modifier within one conjunct (giving the split reading). That is, the variation between the split and joint reading for coordinated adjectives phrases is due to a structural ambiguity. This is accomplished by the phrase-structure rules in (34) and (35), which are the basic templates for all languages (discussed here); as we will see, each language varies these templates slightly.[9]

(34) Phrase-structure rule for adjective coordination, joint reading:

AdjP	→	AdjP^{+}	Cnj	AdjP
		$\downarrow \in \uparrow$	$\uparrow = \downarrow$	$\downarrow \in \uparrow$

Rule (34) for the joint reading is the familiar LFG rule for coordination, creating a set of f-structures: here, a set of AdjP f-structures which will appear in the adjunct (ADJ) set of the modified noun phrase, as shown in (36).

[7] Exceptionless nondistributivity of CONCORD would potentially mean that INDEX and CONCORD were indistinguishable, i.e. in such a language there would effectively be no distinction between these two types of features, except for cases where there are mismatches between CONCORD and INDEX, as found e.g. in Slavic (Wechsler and Zlatić 2003, Hristov 2012). In all of the languages discussed in this paper, CONCORD is distributive in at least some constructions.

[8] Przepiórkowski and Patejuk (2012) have proposed treating distributivity as a property of statements rather than features themselves. However, for our purposes, as we shall see below, there is no need to make any modifications to the LFG formalism.

[9] We use the following abbreviations in phrase-structure rules:

i. CNUM ≡ CONCORD NUM ii. INUM ≡ INDEX NUM

(35) Phrase-structure rule for adjective coordination, split reading:

AdjP	→	AdjP^{+}	Cnj	AdjP
		$\downarrow \in (\%C\ \text{ADJ})$	$\uparrow = \downarrow$	$\downarrow \in (\%C\ \text{ADJ})$
		$\%C \in \uparrow$	$(\uparrow \text{INUM}) = \text{PL}$	$\%C \in \uparrow$

Rule (35) for the split reading creates a set of "incomplete" (PRED-less) NPs with the adjective occupying the ADJUNCT set. This is accomplished by use of the local name %C, which appears in each daughter category in the rule; recall that the scope of a local name is limited to the annotations on the daughter category in which it appears, so that %C in each daughter category refers to a separate f-structure (Dalrymple 2001, Crouch et al. 2008). The annotations require each daughter AdjP to introduce a member %C of a set representing a coordinate structure of the type expected for coordinated nouns, where each conjunct has an adjunct (ADJ) set containing a modifier, as shown in (37). The PRED feature is a distributive feature, and so in the latter case, the noun's PRED value distributed into this set. This gives us the following f-structures for the two kinds of coordination:

(36) Joint reading (rule 34):

$$\begin{bmatrix} \textit{noun} & \\ \text{ADJ} & \left\{ \begin{bmatrix} \text{CONJ} & \text{AND} \\ \left\{ \left[\textit{adj1} \right], \left[\textit{adj2} \right] \right\} & \end{bmatrix} \right\} \end{bmatrix}$$

(37) Split reading (rule 35):

$$\left\{ \begin{bmatrix} \text{CONJ} & \text{AND} \\ \left\{ \begin{bmatrix} \textit{noun} & \\ \text{ADJ} & \left\{ \left[\textit{adj1} \right] \right\} \end{bmatrix}, \begin{bmatrix} \textit{noun} & \\ \text{ADJ} & \left\{ \left[\textit{adj2} \right] \right\} \end{bmatrix} \right\} & \end{bmatrix} \right\}$$

The split reading for 'red and white flags' (in the 'exactly two' reading) thus involves an f-structure roughly equivalent to 'the red flag and the white flag'. Clearly, the distribution of the PRED value in (37) cannot be achieved using the standard adjunction rule, since the coordinated adjective phrase in the split reading introduces an f-structure resembling the structure for coordinated nouns; hence, unification is required. Hence the AdjP adjunction rule must involve an option for the f-structure for the AdjP to be a co-head:

(38)

N′	→	AdjP	N′
		$\{\downarrow \in (\uparrow \text{ADJ}) \mid \uparrow = \downarrow\}$	$\uparrow = \downarrow$

Note that the co-head possibility for the AdjP can only be used in conjunction with the rule for coordinated AdjPs with a split reading; its use with a simple AdjP is ruled out, since a PRED clash would result.

4.1 Hindi

We assume that CONCORD is a distributive feature in Hindi. For the purposes of this paper, we assume that attributive adjectives head AdjPs in Hindi.[10] Since Hindi does not allow the coordination of singular adjectives to have a split reading, we have to modify the rule in (35) to enforce plural concord:[11]

(39) Phrase-structure rule for adjective coordination, split reading (Hindi):

AdjP	→	AdjP$^+$	Cnj	AdjP
		$\downarrow \in (\%C \text{ ADJ})$	$\uparrow = \downarrow$	$\downarrow \in (\%C \text{ ADJ})$
		$\%C \in \uparrow$	$(\uparrow \text{ INUM}) = \text{PL}$	$\%C \in \uparrow$
			$(\uparrow \text{ CNUM}) = \text{PL}$	

For the joint reading of the Hindi phrase in (22), the resulting f-structure is as in (40), whereas for the same phrase with a split reading the f-structure will be as in (41).

(40) Joint reading (rule 34):

$$\begin{bmatrix} \text{PRED} & \text{'flag'} \\ \text{SPEC} & \begin{bmatrix} \text{PRED} & \text{'these'} \end{bmatrix} \\ \text{CONCORD} & \begin{bmatrix} \text{NUM} & \text{PL} \end{bmatrix} \\ \text{INDEX} & \begin{bmatrix} \text{NUM} & \text{PL} \end{bmatrix} \\ \text{ADJ} & \left\{ \begin{bmatrix} \text{CONJ} & \text{AND} \\ \left\{ \begin{bmatrix} \text{PRED} & \text{'yellow'} \end{bmatrix} \right. \\ \left. \begin{bmatrix} \text{PRED} & \text{'green'} \end{bmatrix} \right\} \end{bmatrix} \right\} \end{bmatrix}$$

(41) Split reading (rule 39):

$$\begin{bmatrix} \text{SPEC} & \begin{bmatrix} \text{PRED} & \text{'these'} \end{bmatrix} \\ \text{INDEX} & \begin{bmatrix} \text{NUM} & \text{PL} \end{bmatrix} \\ \text{CONJ} & \text{AND} \\ \left\{ \begin{bmatrix} \text{PRED} & \text{'flag'} \\ \text{CONCORD} & \begin{bmatrix} \text{NUM} & \text{PL} \end{bmatrix} \\ \text{ADJ} & \left\{ \begin{bmatrix} \text{PRED} & \text{'yellow'} \end{bmatrix} \right\} \end{bmatrix} \right. \\ \left. \begin{bmatrix} \text{PRED} & \text{'flag'} \\ \text{CONCORD} & \begin{bmatrix} \text{NUM} & \text{PL} \end{bmatrix} \\ \text{ADJ} & \left\{ \begin{bmatrix} \text{PRED} & \text{'green'} \end{bmatrix} \right\} \end{bmatrix} \right\} \end{bmatrix}$$

Here, the coordinated AdjP is either a functional co-head with the noun (as in 35), or specified by the phrase structure rule as appearing in the ADJ set (as in 34).

[10] The English facts are equivalent to the Hindi, except for two differences: English adjectives lack number marking, and attributive adjectives in English are non-projecting (Sadler and Arnold 1994, Toivonen 2003, Arnold and Sadler 2013). The only differences required for English therefore relate to the categories involved: (34) and (35) will involve coordination of the non-projecting category $\widehat{\text{Adj}}$, not AdjP, and (38) will involve adjunction of $\widehat{\text{Adj}}$ to N^0. In other respects, in particular the annotations, the rules will be identical.

[11] In this and following rules, the language-specific annotations are given in plain face, the common annotations (those in 34 or 35) are given in grey.

4.2 Italian

For Italian, the rule for the joint reading (34) is unchanged. For the phrase in (1) with the joint reading, the resulting f-structure will be as in (42).

(42) Joint reading (rule 34):

$$\begin{bmatrix} \text{PRED} & \text{'station'} \\ \text{CONCORD} & \begin{bmatrix} \text{NUM} & \text{SG} \end{bmatrix} \\ \text{INDEX} & \begin{bmatrix} \text{NUM} & \text{SG} \end{bmatrix} \\ \text{ADJ} & \left\{ \begin{bmatrix} \text{CONJ} & \text{AND} \\ \left\{ \begin{bmatrix} \text{PRED} & \text{'old'} \end{bmatrix} \right. \\ \left. \begin{bmatrix} \text{PRED} & \text{'small'} \end{bmatrix} \right\} \end{bmatrix} \right\} \end{bmatrix}$$

Since Italian allows 'resolving' agreement, the crucial difference between Italian and Hindi is that CONCORD is nondistributive in Italian when AdjPs are coordinated. This means that agreement is not enforced between the noun and each adjective conjunct; rather, the plural CONCORD and INDEX features of the noun match the resolved plural CONCORD and INDEX features of the coordinated AdjP. With this proviso, the Hindi variant of the rule for the split reading (39) can be used without modification for the pattern illustrated in (5) with singular coordinated adjectives and a plural noun. Since the resolving pattern in Italian only occurs with postnominal AdjPs, we need separate coordination and adjunction rules for them:

(43) Phrase-structure rule for Italian AdjP adjunction:

$$\begin{array}{lcll} \text{N}' & \rightarrow & \text{N}' & \text{AdjP} \\ & & \uparrow{=}\downarrow & \{\downarrow \in (\uparrow \text{ADJ}) \mid \uparrow{=}\downarrow\} \end{array}$$

(44) Phrase-structure rule for Italian AdjP coordination, split reading (same as 39):

$$\begin{array}{lcccc} \text{AdjP} & \rightarrow & \text{AdjP}^{+} & \text{Cnj} & \text{AdjP} \\ & & \downarrow \in (\%C \text{ ADJ}) & \uparrow{=}\downarrow & \downarrow \in (\%C \text{ ADJ}) \\ & & \%C \in \uparrow & (\uparrow \text{INUM}) = \text{PL} & \%C \in \uparrow \\ & & & (\uparrow \text{CNUM}) = \text{PL} & \end{array}$$

This rule allows the coordinate adjectives in the "resolving" pattern to have different number features, which is indeed allowed in Italian:

(45) *tulipani* *rosso* *e* *bianchi*
tulips.PL red.SG and white.PL
'red and white tulips' (Italian, caption of a picture showing one red tulip and two white tulips[12])

[12] https://www.flickr.com/photos/orsorama/8704984416/

For the phrase in (5), which has only a split reading, the resulting f-structure will be as in (46).

(46) Split reading (rule 35):

$$\begin{bmatrix}
\textsc{concord} & \begin{bmatrix}\textsc{num} & \textsc{pl}\end{bmatrix} \\
\textsc{index} & \begin{bmatrix}\textsc{num} & \textsc{pl}\end{bmatrix} \\
\textsc{conj} & \textsc{and} \\
\left\{\begin{bmatrix}
\textsc{pred} & \text{`flag'} \\
\textsc{concord} & \begin{bmatrix}\textsc{num} & \textsc{sg}\end{bmatrix} \\
\textsc{adj} & \left\{\begin{bmatrix}\textsc{pred} & \text{`red'}\end{bmatrix}\right\}
\end{bmatrix}\right. \\
\left.\begin{bmatrix}
\textsc{pred} & \text{`flag'} \\
\textsc{concord} & \begin{bmatrix}\textsc{num} & \textsc{sg}\end{bmatrix} \\
\textsc{adj} & \left\{\begin{bmatrix}\textsc{pred} & \text{`white'}\end{bmatrix}\right\}
\end{bmatrix}\right\}
\end{bmatrix}$$

For the non-resolving agreement pattern in Italian (6–7), in which the noun as well as the adjectives are singular, we require a different adjunction rule (for prenominal $\widehat{\text{Adj}}$s) as well as a special coordination rule for $\widehat{\text{Adj}}$s:

(47) Phrase-structure rule for Italian $\widehat{\text{Adj}}$ adjunction:

$$\begin{array}{lcll}
\text{N}^0 & \rightarrow & \widehat{\text{Adj}} & \text{N}^0 \\
 & & \{\downarrow\in(\uparrow \textsc{adj}) \mid \uparrow=\downarrow\} & \uparrow=\downarrow
\end{array}$$

(48) Phrase-structure rule for Italian $\widehat{\text{Adj}}$ coordination, joint reading:

$$\begin{array}{lcllll}
\widehat{\text{Adj}} & \rightarrow & \widehat{\text{Adj}}^{+} & \text{Cnj} & \widehat{\text{Adj}} \\
 & & \downarrow\in\uparrow & \uparrow=\downarrow & \downarrow\in\uparrow
\end{array}$$

(49) Phrase-structure rule for Italian $\widehat{\text{Adj}}$ coordination, split reading:

$$\begin{array}{lcccc}
\widehat{\text{Adj}} & \rightarrow & \widehat{\text{Adj}}^{+} & \text{Cnj} & \widehat{\text{Adj}} \\
 & & \downarrow\in(\%C\ \textsc{adj}) & \uparrow=\downarrow & \downarrow\in(\%C\ \textsc{adj}) \\
 & & \%C\in\uparrow & (\uparrow \textsc{inum})=\textsc{pl} & \%C\in\uparrow \\
 & & (\uparrow \textsc{cnum})=(\downarrow \textsc{cnum}) & & (\uparrow \textsc{cnum})=(\downarrow \textsc{cnum})
\end{array}$$

The annotations under the coordinated phrases in (49) differ from the rule in (35) in two respects. First, the Cnj node is not annotated with a value for the CNUM feature; this allows either singular or plural adjectives to participate in this construction, since the CNUM value of the coordinate adjective phrase is not constrained. Second, each AdjP conjunct is annotated with the requirement for its CNUM value to match the nondistributive CNUM value of the coordinated AdjP as a whole. This enforces the requirement for the conjuncts to have uniform number, either singular (as in example 6) or plural (as in example 11).

We also require that the lexical entries of singular nouns in Italian contain the following specifications:

(50) biblioteca N
$(\uparrow \text{PRED}) = \text{`library'}$
$(\uparrow \text{CNUM}) = \text{SG}$
$\{\ (\uparrow \text{INUM}) = \text{SG} \mid (\uparrow \text{INUM}) =_c \text{PL}\ \}$

The constraining equation in the last line permits a singular noun to function as the head of a phrase with INDEX NUM = PL, only if this feature is specified elsewhere, i.e. if it is specified in the phrase structure rules.[13] We will therefore get the f-structure in (51) for the noun phrase in (6).

(51) Split reading (rule 49):

$$\left[\begin{array}{ll}
\text{CONCORD} & \left[\text{NUM} \quad \text{SG}\right] \\
\text{INDEX} & \left[\text{NUM} \quad \text{PL}\right] \\
\left\{\left[\begin{array}{ll}
\text{PRED} & \text{`library'} \\
\text{CONCORD} & \left[\text{NUM} \quad \text{SG}\right] \\
\text{ADJ} & \left\{\left[\text{PRED} \quad \text{`old'}\right]\right\}
\end{array}\right]\right. & \\
\left.\left[\begin{array}{ll}
\text{PRED} & \text{`library'} \\
\text{CONCORD} & \left[\text{NUM} \quad \text{SG}\right] \\
\text{ADJ} & \left\{\left[\text{PRED} \quad \text{`new'}\right]\right\}
\end{array}\right]\right\} &
\end{array}\right]$$

4.3 Russian

The only difference between Russian and Italian is that Russian allows both the resolving and non-resolving patterns with any adjective regardless of its lexical class or syntactic position. This may be modeled by including two rules for AdjP coordination in Russian grammar: the one in (44) and one analogous to (49), but for AdjPs. We may also generalize over these readings by introducing a unified rule like the following:

[13] In this way, singular number in Italian is in some sense an unmarked number: it reflects SG by default, but can reflect PL if externally specified.

(52) Phrase-structure rule for Russian AdjP coordination, split reading:

AdjP	→	AdjP^{+}	Cnj	AdjP
		$\downarrow \in (\%C$ ADJ)	$\uparrow = \downarrow$	$\downarrow \in (\%C$ ADJ)
		$\%C \in \uparrow$	(↑ INUM) = PL	$\%C \in \uparrow$
		{(↑ CNUM) = PL |		{(↑ CNUM) = PL |
		(↑ CNUM) = (↓ CNUM)}		(↑ CNUM) = (↓ CNUM)}

5 Further issues

In this paper, we have addressed only adjective coordination, but in Russian the same, or similar, effects are observed with other NP subconstituents:

(53) *Pasportistka 12-go otdelenija milicii dvaždy*
passport.officer of.twelfth station of.police twice
podyšav na štamp «Propisan postojanno», ottisnula ego
having.breathed on stamp registered permanently imprinted it
*na **pasportax moëm i ženy***
on passport.PL my.M.PREP.SG and wife.GEN.SG
'The passport officer, having breathed twice on the stamp "Permanently registered", imprinted it on **me and my wife's passports**.' (Russian, RNC: Vladimir Vojnovič. Ivan'kiada, ili rasskaz o vselenii pisatelja Vojnoviča v novuju kvariru, 1976)

In (53), a plural noun is modifed by an 'unlike' coordination of a singular possessive adjective and a genitive case noun. Thus the analysis must be extended to cover at least case-marked NPs, and possibly also other phrase types, such as PPs. This would also require a theory of coordination of unlikes. The situation is especially complicated by the fact that in (53) we are dealing with a so-called possessive adjective 'my', an element which is syntactically and morphologically an adjective, but which is functionally equivalent to a genitive dependent, i.e. to a POSS or COMP, but not an ADJ.

A potential problem with the present account is that the stipulation of two different coordination rules and two different adjunction rules significantly increases grammatical complexity. While such a solution seems to be unavoidable in the current LFG architecture, it remains to be seen whether this kind of ambiguity of coordinating constructions is necessary elsewhere in the grammar.

6 Conclusion

In this paper, we have provided an analysis that adequately describes the observed effects of agreement resolution in adjective coordination. While resolving agreement is a problem for theories of agreement which predict that an adjective should not be able to show different agreement features from its controller, the analysis

faces few difficulties in a symmetric theory of agreement, such as is standard in LFG. The LFG formalization that we proposed is based on the distinction between CONCORD and INDEX features to differentiate between NP-internal and clause-level agreement. Mismatches within NPs (i.e. plural marking of the head noun combined with coordinated singular adjectives) are described by assuming that the "resolving" agreement type involves a special, "unificational", rule of coordination, where it is not adjectives themselves that are coordinated, but f-structures lacking PRED features which contain the adjectives as their adjuncts. This requires the introduction of an additional rule of adjective coordination in addition to the standard set-membership one, and of an additional annotation on AdjP which allows it to act as a co-head. As noted above, some of the constructions we analyze in this paper have formerly been described in terms of ellipsis. In this light it is interesting to note that our analysis involves ellipsis-like effects at f-structure without involving any deletion *per se*. It remains for future work to establish whether the distinction between two kinds of coordination that we have introduced is useful for other similar constructions.

Another important distinction on which our analysis depends is the distinction between distributive and non-distributive features. We have demonstrated that the availability of the resolving and non-resolving agreement types shows much cross-linguistic variation. In the former type, CONCORD behaves like a non-distributive feature, while in the latter, it behaves like a distributive one. However, simply assuming that CONCORD can be non-distributive does not solve the problem, because this would run counter to other constructions where no such effects are observed. Furthermore, within some languages, like Italian and Russian, both the distributive and the non-distributive agreement types are allowed. We claim that the most adequate solution to this issue is to assume that distributivity can be construction-specific. This can be modeled without modifying the basic LFG architecture by simulating distributivity in given constructions through annotations on individual conjuncts.

References

ARNOLD, DOUG and LOUISA SADLER (2013). 'Displaced Dependent Constructions'. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG13 Conference*, Stanford, CA: CSLI Publications, pp. 48–68.

BARONI, MARCO, SILVIA BERNARDINI, FEDERICA COMASTRI, LORENZO PICCIONI, ALESSANDRA VOLPI, GUY ASTON, and MARCO MAZZOLENI (2004). 'Introducing the La Repubblica corpus: A large, annotated, TEI(XML)-compliant corpus of newspaper Italian'. In *Proceedings of LREC 2004*.

BOSQUE, IGNACIO (2006). 'Coordinated adjectives and the interpretation of number features'. In Laura Brugè (ed.), *Studies in Spanish Syntax*, Venice: Libreria Editrice Cafoscarina, pp. 47–60.

CORBETT, GREVILLE G. (2006). *Agreement*. Cambridge University Press.

CROUCH, DICK, MARY DALRYMPLE, RON KAPLAN, TRACY KING, JOHN MAXWELL, and PAULA NEWMAN (2008). *XLE Documentation*. Palo Alto Research Center, Palo Alto, CA.

DALRYMPLE, MARY (2001). *Lexical Functional Grammar*, volume 34 of *Syntax and Semantics*. New York: Academic Press.

DALRYMPLE, MARY and RONALD M. KAPLAN (2000). 'Feature Indeterminacy and Feature Resolution'. *Language 76*(4), pp. 759–798.

DALRYMPLE, MARY and IRINA NIKOLAEVA (2006). 'Syntax of natural and accidental coordination: Evidence from agreement'. *Language 82*(4), pp. 824–849.

HEYCOCK, CAROLINE and ROBERTO ZAMPARELLI (2005). 'Friends and Colleagues: Plurality, Coordination, and the Structure of DP'. *Natural Language Semantics 13*, pp. 201–270.

HRISTOV, BOZHIL (2012). *Agreement, Case Assignment, and Nominal Coordination*. DPhil thesis, University of Oxford.

IOMDIN, LEONID L. (1990). *Avtomatičeskaja obrabotka teksta na estestvennom jazyke: model' soglasovanija [Automatic natural language processing: The model of agreement]*. Moscow: Nauka.

KIBRIK, ANDREJ A. (2011). *Reference in Discourse*. Oxford University Press.

KING, TRACY HOLLOWAY and MARY DALRYMPLE (2004). 'Determiner agreement and noun conjunction'. *Journal of Linguistics 40*, pp. 69–104.

KODZASOV, SANDRO V. (1987). 'Čislo v sočinitel'nyx konstrukcijax [Number in coordinating constructions]'. In Alexander E. Kibrik and Aleksandr S. Narin'jani (eds.), *Modelirovanie jazykovoj dejatel'nosti v intellektual'nyx sistemax*, Moscow: Nauka, pp. 201–219.

MAXWELL, JOHN T., III and CHRISTOPHER D. MANNING (1996). 'A theory of non-constituent coordination based on finite state rules'. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG96 Conference*, Stanford, CA: CSLI Publications.

PRZEPIÓRKOWSKI, ADAM and AGNIESZKA PATEJUK (2012). 'On Case Assignment and the Coordination of Unlikes: The Limits of Distributive Features'. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG12 Conference*, Stanford, CA: CSLI Publications, pp. 479–489.

SADLER, LOUISA and DOUG ARNOLD (1994). 'Prenominal Adjectives and the Phrasal/Lexical Distinction'. *Journal of Linguistics 30*, pp. 187–226.

STEELE, SUSAN (1978). 'Word order variation: a typological study'. In Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik (eds.), *Universals of Human Language IV: Syntax*, Stanford University Press, pp. 585–623.

TOIVONEN, IDA (2003). *Non-Projecting Words: A Case Study of Swedish Verbal Particles*. Dordrecht: Kluwer.

WÄLCHLI, BERNHARD (2005). *Co-Compounds and Natural Coordination*. Oxford: Oxford University Press.

WECHSLER, STEPHEN and LARISA ZLATIĆ (2003). *The many faces of agreement*. Stanford, CA: CSLI Publications.

MASDARS AND MIXED CATEGORY CONSTRUCTIONS

Kersti Börjars, Safiah Madkhali
and John Payne

The University of Manchester

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King (Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

In Arabic grammar the term MASDAR is applied to a variety of noun forms derived from verb and other stems according to a set of partially regular but largely irregular morphological patterns (Ryding 2005). Three deverbal masdar forms allow complex event interpretations in which the verbal argument structure is inherited, but at the same time, they display a range of nominal properties. In both transformational and LFG treatments, these phrases have been assumed to have both verbal and nominal syntax. We propose instead that verbal functions be permitted inside what is categorially an NP from top to bottom.

1 Introduction

In Arabic grammar, the term MASDAR is applied to a variety of noun forms derived from verb and other stems according to a set of partially regular but largely irregular morphological patterns (Ryding 2005). The three deverbal masdar forms which potentially allow complex event interpretations are: (i) the basic masdar (BM), which is the most productive and heterogeneous form and has the widest distribution (e.g. *ʔaṯāba* 'reward' ~ *ʔiṯāba*(*t*) (reward.BM) 'rewarding'); (ii) the *mim* masdar (MM), which is less productive and distributionally more restricted, and characteristically prefixed by *m*V- (e.g. *maṯōba*(*t*) (reward.MM) 'rewarding'); and (iii) the non-stem-derived masdar (NSDM), whose formation is essentially unpredictable and does not include all the consonants of the root (e.g. *ṯawāb* (reward.NSDM) 'rewarding').

There are two different constructions involving the masdar which have mixed category properties, we will refer to the two types as Masdar Mixed Construction A (MMC A) and Masdar Mixed Construction B (MMC B). All of the masdar forms mentioned above occur in both MMC A and MMC B.

2 Masdar Mixed Construction A

2.1 Properties

We will illustrate here with the BM, but the MM and NSDM show an identical apparent mixture of nominal and verbal characteristics. Like verbs, and unlike nominals generally, masdars derived from transitive verbs characteristically take an accusative object in MMC A as illustrated in the basic monotransitive example in (1).

(1) **ʔakl-u** **l-walad-i** it-tufāhat-a
eat.BM-NOM DEF-boy-GEN DEF-apple.ACC
'the boy's eating the apple'

A further verbal property is that the masdars in the MMC A construction permit **only** adverbial modification, as in (2a), involving a monotransitive verb with a clitic genitive pronoun. The adverb *muʔakk̲aran* 'recently' must follow the object rather than precede it as in (2b), and cannot be substituted by the corresponding adjective in either position (see 2c and 2d):[1]

(2) a. **tansīq -u=hā** iz-zuhōr-a muʔakk̲aran
arrange.BM-NOM=3FS.GEN DEF-flowers-ACC recently
'her arranging the flowers recently'

b. ***tansīq-u=hā** muʔakk̲aran iz-zuhōr-a
arrange.BM-NOM=3FS.GEN recently DEF-flowers-ACC

c. ***tansīq-u=hā** iz-zuhōr-a ʔal-ʔakk̲i:r-u
arrange.BM-NOM =3FS.GEN DEF-flowers-ACC DEF-last-NOM

d. ***tansīq-u=hā** ʔal-ʔakk̲i:r -u iz-zuhōr-a
arrange.BM-NOM =3FS.GEN DEF-last-NOM DEF-flowers-ACC

As shown in (3), the masdar can also inherit two accusative object arguments.

(3) **taslīm-a=hā** il-muwaẓaf-īna
hand.BM-ACC=3FS.GEN DEF-employee-PL.ACC
rawātib-a=hum fawran wa bidōni taʔxīr
salaries-ACC=3PL.GEN immediately and without delay
'its [the company's] handing the employees their salaries immediately and without delay'

[1] In Arabic, the vast majority of adverbs are derived from adjectives, in which case they assume the accusative form of the corresponding adjective. Unlike adjectives, they are invariant, and show no agreement in case, gender, number or definiteness. In some cases, as in the examples in (2), the relationship between adjective and adverb is not completely transparent. The adverb *muʔakk̲aran* 'recently' has no direct adjectival source. The adjective *ʔakk̲i:r-* 'last' has the same root, but has a different meaning and form. Also, just a handful of adverbs are morphologically simple, e.g. *jiddan* 'very'.

Example (3) also contains adverbial modification, the adverbial in this case represented by a coordinated adverb and adjunct PP. Adjectival modification would again be impossible.

These essentially verbal properties (accusative objects and inability to take adjectival modification) are combined with a number of nominal characteristics. The whole construction has the external distribution of an NP.[2] Another salient nominal property of the masdar itself is its ability to be marked for case, e.g. nominative in (1) and (2a) and accusative in (3). Furthermore, the masdar and the immediately following genitive NP (either a clitic pronoun or a full NP) form the tight-knit CONSTRUCT STATE (CS) construction typical of basic possessive NPs in Arabic. In the monotransitive examples in (1) and (2), and the ditransitive example in (3), the genitive NP represents the highest, typically agent, argument, and the accusative NP or NPs represent lower arguments, typically theme or recipient.

The CS construction found in (1), (2) and (3) is the same construction that is used more generally to indicate possessor relations in the NP, as exemplified in (4).

(4) tufāhat-u l-walad-i
apple-NOM DEF-boy-GEN
'the apple of the boy'

It is also a component of MMC B, as will be shown in Section 3 below.

Some crucial characteristic properties of the CS construction are shared by all the MMC A examples in (1), (2) and (3). Nothing can intervene between the head noun and following genitive NP. In both construction types the head noun itself is not marked definite, even though the whole NP is definite. This property of the CS construction, DEFINITENESS INHERITANCE, has been widely discussed and has received varying analyses in the literature. An LFG analysis of analogous CS constructions in Hebrew is for example to be found in Falk (2001). See also Danon (2008), and the discussion of Arabic CS constructions with adjectival heads in Alsharifi & Sadler (2009). Definite inheritance bears some resemblance to the definiteness effect observed in many languages, where the mere presence of a possessive determiner imposes a definite interpretation on the NP as a whole (see Haspelmath (1999) for a typological discussion). It differs however in that only NPs whose possessive NPs are themselves definite receive a definite interpretation.

[2] We assume that full nominal arguments are of the category NP — rather than for instance DP — in Arabic, but nothing hinges on this assumption. See (7) for assumed annotated c-structure rules for Arabic noun phrases.

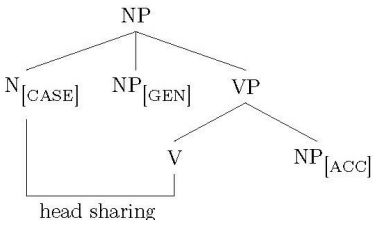
2.2 Previous analyses

Early analyses of the MMC A construction in Arabic took a transformational form involving movement of a verbal head to a higher nominal head position, e.g. Fassi-Fehri (1993). Such an analysis would in principle not be available within a lexicalist architecture such as LFG. MMC A rather seems to fall squarely within the ambit of LFG head-sharing analyses, and indeed is placed there (though without detailed analysis) by Bresnan (1997). Al-Sharif (2014) makes the same suggestion.

There are some issues, though. In the classic head-sharing analysis of mixed category event nominalizations such as that applied to the Italian *infinito sostantivato* construction in Bresnan (1997), there are, as in the earlier movement analysis, two structural heads of differing categories: a V and an N. The nominalized form of the verb (in the Italian example an infinitive form) is the V head of a VP which also houses any core NP objects and accompanying adverbials. This VP is then the sister of an N which projects to the NP level and houses any properly nominal constituents of the construction such as determiners and adjectives. Technically, the higher N functions as an extended head of the lower V, and it is the lower V which provides the PRED value of the corresponding f-structure.

An attempt to apply an analysis of this kind to a monotransitive MMC A might schematically look like (5):

(5)

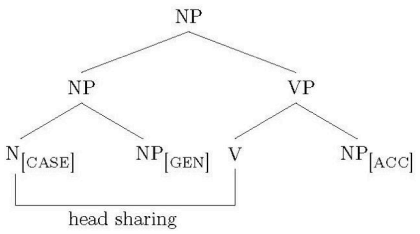


Note that the masdar in MMC A, unlike the Italian infinitive, would have to sit in the higher N position, where it would provide the PRED value of the corresponding f-structure. This is firstly because it is separated from the accusative object by the genitive NP, and secondly because it is clearly a noun. For example, it displays characteristically nominal properties such as being marked for case. This is in itself unproblematic: a configuration in which the extended N head must be analysed as lexically filled in a head-sharing analysis rather than the lower V head is found in the agent nominalisations of Gĭkŭyŭ (Bresnan & Mugane 2006). The issue with (5) rather rests in the fact that as it stands it fails to capture the tight-knit nature of the CS construction consisting of the head noun and following genitive NP. In particular, in the basic CS construction as illustrated in (4), and indeed in the CS construction found in MMC B, the combination of the head noun and following genitive NP forms a

constituent. The constituency of the CS in MMC B is shown in example (13) below.

An alternative to (5) which preserves the notion of head-sharing, but treats the CS as a constituent, might schematically be represented as (6):

(6)



In this analysis, the masdar would similarly sit in the N head position, but the VP housing the accusative object (and any adverbials) would sit in an adjunct position (here represented as sister and daughter of NP). This is analogous to the analysis of Dagaare agent nominalizations in Bresnan (1997), where the posited VP housing objects and adverbials must likewise be separated from the head noun by an overtly nominal constituent (in the Dagaare case a modifying adjective). An analysis of this kind raises both technical and conceptual issues. Firstly, it is difficult to reconcile the notion of extended head with this configuration, where the N does not stand in a sister relation to the VP. Secondly, the adjunct function of the VP does not sit easily with the potential obligatoriness of the object in the construction.

2.3 Proposed analysis for MMC A

Instead therefore of analysing the Arabic masdar constructions as head-sharing constructions in the classic sense, we propose here an analysis in which these constructions are purely nominal from top to bottom. The job of permitting object functions within the NP will fall to f-structure. The masdar forms in question will inherit their functional structure directly from the corresponding verbs, and this will result in a principled extension to the realisational potential of nominal forms denoting complex events, rather than entailing the presence of a concomitant syntactic VP.

This analysis in some ways reflects the HPSG analysis of English verbal gerunds in Malouf (2000). Malouf observes that the oddity of verbal gerunds lies in the fact that they are noun-like in terns of their category, but verb-like in terms of their selectional properties. This is translated in Malouf's analysis into the notion that a verbal gerund head can subcategorize for NP complements, while at the same time preserving nominal characteristics. In LFG terms, subcategorisation principles convert into f-structure principles of argument

selection. Unlike Malouf, however, we will not treat nominalized heads such as masdars as belonging to a separate lexical category (verbal gerund) which inherits its properties from both nouns and verbs. In our analysis, the forms in question are unequivocally nouns.

To be precise, in MMC A the tight-knit CS constituent will be extended to include accusative objects. This captures a generalization that covers all the core arguments: they are sisters of the masdar. It also reflects the fact that these accusative objects are, similarly to the genitive NP representing the highest argument, tightly bound to the masdar. Just as nothing can intervene between a masdar in the CS and the following genitive NP, neither can anything intervene between the genitive NP and a following object NP, or indeed between two object NPs in case of ditransitive forms. To this constituent consisting of the masdar and its core arguments obliques and/or adjuncts can be added in free order.

The annotated c-structure rules which license the proposed structure for MMC A are given in (7):

(7) (i) NP → N NP NP NP
↑=↓ (↓CASE)=gen (↓CASE)=acc (↓CASE)=acc
(↑SUBJ)=↓ (↑OBJ)=↓ (↑OBJ_θ)=↓

(ii) NP → NP PP
↑=↓ (↑OBL)=↓

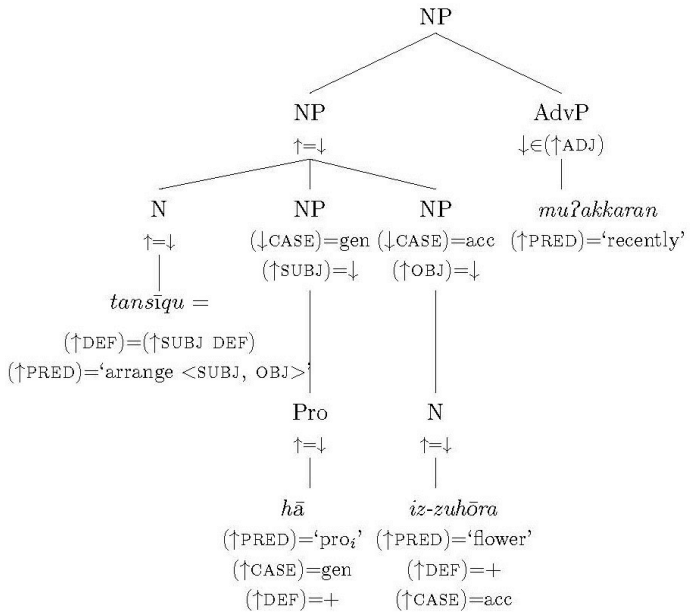
(iii) NP → NP XP
↑=↓ ↓ ∈ (↑ADJ)

As usual, these rules are to be construed as maximal. In particular, the presence of any of the dependent NPs in (7i) will be licensed by the requirements of the head noun, and in particular its argument structure.[3]

A proposed analysis for (2a) is then provided in (8).

[3] A complication here is that the genitive NP in MMC A is optional. The first argument must then be interpreted as a pronominal subject via anaphoric control.

(8)



It will be noted that the definiteness inheritance property of the CS emerges in this analysis from the lexical entry of the head noun, which takes a distinctive form marked neither by the definite prefix *il-* and its variants, nor by the indefinite suffix *–n*. Nouns in this distinctive form bear the annotation (↑DEF) = (↑SUBJ DEF) which will force them to co-occur with a SUBJ, and the definiteness of the NP as a whole will be inherited from this SUBJ. In spirit this follows Alsharifi and Sadler's (2009) lexical treatment of the adjectival construct state in Arabic, though the details of the nominal and adjectival constructions are different.

We assume that the verbal argument structure is inherited in its entirety by the masdar and is as illustrated in (9).

(9) 'tansīq- <(arg$_1$, arg$_2$) >'
−o −r
| |
SUBJ OBJ

The double object argument structure of the masdar in (3) is also directly inherited:

(10) 'taslīm-- < (arg$_1$, arg$_2$ arg$_3$)>'
−o −r +o
| | |
SUBJ OBJ OBJ$_\Theta$

Following the conventions proposed in Kibort (2014), we label the arguments as arg_1, arg_2 and arg_3. The first argument will be realized as a SUBJ, which inside a nominal will be marked by genitive case. The second argument will be an OBJ, and marked accusative. The third argument in the double object construction is an OBJ_Θ and likewise accusative.

3 Masdar Mixed Construction B

All the masdar forms in question permit an alternative construction to MMC A in which a second argument is expressed not by the accusative case, but rather by a PP headed by the preposition *li-* 'of/to'. Also, unlike in MMC A, adjectival modification is permitted (Bardeas 2009: 257). Example (11a) is an instance of MMC B:

(11) a. ʔakl-u l-walad-i as-sarīʕ-u li-t-tufāhat-i
eat.BM-NOM DEF-boy-GEN DEF-fast.NOM of-DEF-apple-GEN
'the boy's fast eating of the apple'

b. *ʔakl-u l-walad-i li-t-tufāhat-i as-sarīʕ-u
eat.BM-NOM DEF-boy-GEN of-DEF-apple-GEN DEF-fast.NOM

As shown in (11b), there is an ordering constraint: any adjective in MMC B must occur immediately adjacent to the CS, i.e. directly following the genitive NP. This constraint on adjective ordering is however a constraint which equally applies to the basic CS construction, and is not something which is special to MMC B. We will therefore not discuss it further here.

Adjectival modification is thus an unequivocal indicator that we are dealing with MMC B, just as the presence of an accusative object is an unequivocal indication of MMC A. It should be noted however that MMC B also allows the possibility of modification by adverbs, as illustrated in (12).

(12) tansīq-u=hā il-mutqan-u li-z-zuhōr-i
arranging.BM-NOM=3FS.GEN DEF-perfect-NOM of-DEF-flowers-GEN
muʔaḵḵaran
recently
'her perfect arranging of the flowers recently.'

It can be seen from this example that the presence of the adjective *mutqan-* 'perfect' does not preclude the simultaneous presence of the adverb *muʔaḵḵaran* 'recently'. Because of the constraint on adjective ordering, the adverb necessarily follows the adjective.

The constituency of the CS unit consisting of the head noun and genitive NP can be neatly demonstrated in MMC B by coordination facts: it is possible

to coordinate instances of the CS and modify this coordinate constituent by a single adjectival modifier, as in (13).

(13) faya**ḍ**ān-u in-nahr-i wa infijār-u
flooding.BM-NOM DEF-river-GEN and exploding.BM-NOM

il-burkāni ʔal- mufāji?-āni fi il-bilād
DEF-volcano.GEN DEF-sudden-NOM.DU in DEf-country
'the sudden flooding of the river and exploding of the volcano in the country'

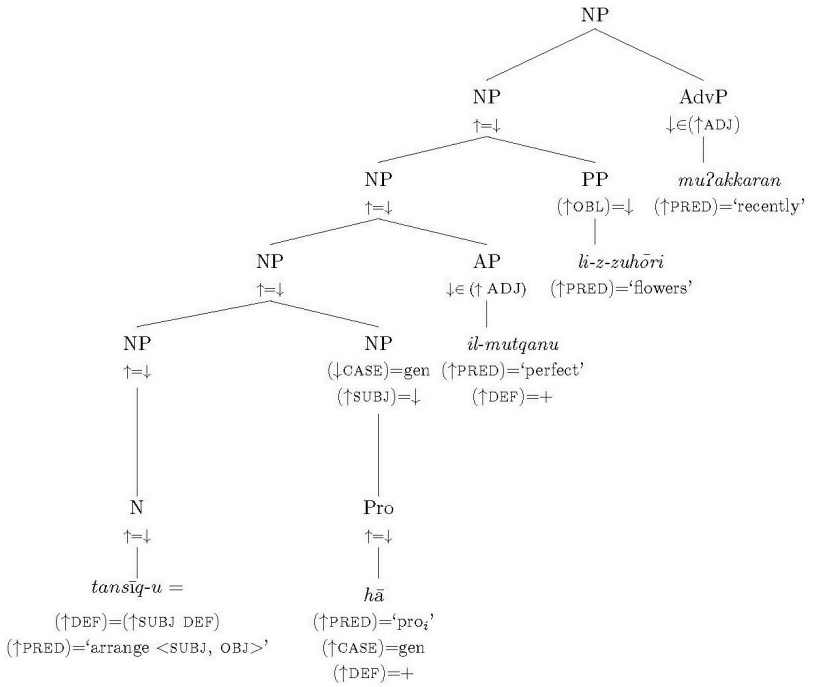
Note that the adjective *mufāji?-* 'sudden' transparently shows by its dual agreement that it must modify both the coordinates. This fact precludes an analysis of the CS construction in which the adjective is internal to the CS, as suggested by Falk (2001) for the analogous construction in Modern Hebrew.

We propose therefore that MMC B requires no addition to the c-structure rules given in (7).[4] The core of the construction is a CS consisting of the masdar and following genitive NP. There are no accusative objects, but rather the possibility of a PP argument as licensed by rule (7ii). Adjectives and adverbial modifiers are both licensed by the adjunct rule (7iii). A proposed analysis of (12) is the provided in (14).[5]

[4] The first NP is omissible and the SUBJ argument can be interpreted (as in MMC A) via anaphoric control. In this case, however, the masdar must be prefixed by the definiteness marker. We leave a treatment of this non-CS variant to future research.

[5] Since it is not directly relevant to our argument, we do not analyse the internal structure of the preposition phrase, and we not provide f-structure features beyond the PRED feature of the noun.

(14)



The PP-argument construction is not an alternative permitted by the corresponding transitive verbal argument structure, and will require the postulation of an alternative more consistently nominal functional structure, with the second argument represented by an oblique. Hence we assume that in this mapping, the second argument is [−o], i.e. not an object, and mapped instead to OBL, as in (15).

(15) 'tansīq- <(arg$_1$, arg$_2$) >'
−o −o
| |
SUBJ OBL

MMC B is more consistently nominal than MMC A in that it allows adjectival modification and disallows accusative objects. However, like MMC A, it permits the presence of adverbs. In this respect Arabic is similar to English, in which the widespread postmodification of nouns by semantically appropriate adverbs has been documented (Payne, Huddleston and Pullum 2010). We assume that the potential for adjectival modification can be based on the semantic structures assigned to mixed and more consistently nominal constructions. See Section 4 below.

Note then that in the case of a masdar construction derived from an intransitive verb, where the genitive NP of the CS represents the sole argument, it is the presence of an adjectival modifier which signals the presence of MMC B. Otherwise, the structural distinction between MMC A and MMC B is neutralized. In particular, the sole presence of modification by an adverb, as in (16), is not sufficient to signal the presence of MMC A.

(16) ʔinhiyār-u il-ḥālat-i iṣ-ṣiḥiyyat-i
deteriorating.BM-NOM DEF-state-GEN DEF-health-GEN

faj ʔatan li il-ʕāhil-i is-suʕōd-i
suddenly of DEF-monarch-GEN DEF-Saudi-GEN
'the deteriorating of the state of health suddenly of the Saudi monarch'

We leave it an open question whether the more mixed MMC A might be considered more marked than the more consistent MMC B. If that were the case, in the absence of a specific indicator of MMC A, i.e. an accusative object, we would have an instance of MMC B.

4 Semantics of Masdar Mixed Constructions

The masdar constructions are transpositions from V to N (cf. Spencer 2005). As such, they reify a verbal structure. However, the two constructions MMC A and MMC B have distinct semantic properties. While MMC B is compatible both with general predicates and temporally specifying predications, MMC A is odd with temporally specifying predications. Compare (17) and (18).

(17) a. laqad fāja ʕa kull-a šāk ṣin [qatl-u
EMPH surprised every-ACC body killing.BM-NOM
il-qāʔid-i il-ẓālimi li-l-junōd-i]
DEF-leader-GEN DEF-unjust.GEN of-DEF-soldiers-GEN
'The unjust leader's killing of the soldiers surprised everyone.'

b. laqad fāja ʕa kull-a šāk ṣin [qatl-u
EMPH surprised every-ACC body killing.BM-NOM
il-qāʔid-i il-ẓālimi il-junōd-a]
DEF-leader-GEN DEF-unjust-GEN DEF-soldiers-ACC
'The unjust leader's killing the soldiers surprised everyone.'

(18) a. laqad ḥada t a [qatl-u il-qāʔid-i
EMPH happened killing.BM-NOM DEF-leader-GEN
il-ẓālimi li-l-junōd-i] fi is-sādisi min Yōlyō
DEF-unjust-GEN of-DEF-soldiers-GEN on DEF-sixth of July
'The unjust leader's killing of the soldiers happened on 6 July.'

b. ʔʔ laqad ḥadaṯa [qatl-u il-qāʔid-i
EMPH happened killing.BM-NOM DEF-leader-GEN
il-ẓālimi il-junōd-a] fi is-sādisi min Yōlyō
DEF-unjust.GEN DEF-soldiers-ACC on DEF-sixth of July
'The unjust leader's killing the soldiers happened on 6 July.'

In (17) we have a general predicate *fājaʕa kull-a šāḵṣin* 'surprised everyone'. This is compatible both with an instance of MMC B, as signaled in (17a) by the presence of the PP argument *li-l-junōd-i* (of-DEF-soldiers-GEN), and an instance of MMC A, as signaled in (17b) by the accusative object *il-junōd-a* (DEF-soldiers-ACC). By contrast, the temporally specifying predicate *ḥadaṯa fi is-sādisi min Yōlyō* 'happened on 6th July' sits readily with MMC B, as in (18a), but is unnatural with MMC A, as in (18b). In Vendler's (1967) terms, predicates like "surprised everyone" are loose containers (they accept both kinds of nominalization), while predicates like "occurred on 6th July" are narrow containers (they accept only the more consistently nominal construction).

This suggests that MMC A and MMC B reify the underlying verbal structures in two different ways. The only detailed formal proposal we are aware of which might capture this difference is Hamm and van Lambalgen (2005), building on earlier work in the formalization of event calculi by Feferman (1984) and Shanahan (1997). Hamm and Lambalgen distinguish between reification of a verbal construction as a FLUENT (which HOLDS AT a particular time), and as an EVENT (which HAPPENS at a particular time). Fluents are initiated and terminated by events. In Hamm and van Lambalgen, this distinction is applied to the English verbal and nominal gerund constructions, but it seems equally applicable to MMC A and MMC B.

In this scenario, MMC A would semantically represent the reification of an underlying proposition (containing a verbal predicate, its arguments and a time variable) as a fluent. The technical instantiation of this reification is abstraction over the time variable, so that a fluent essentially represents the set of times at which the underlying proposition holds. The fluent *the unjust leader's killing the soldiers* would then have the representation in (19a), with *a* being a time variable and *â* the abstraction over it. The relation between a fluent and the underlying proposition is very direct, and in the case of the MMC A in (17a) and (17b) could be represented as in (19b).

(19) a. kill[l, s, â]
b. HoldsAt(kill[l, s, â], t) ↔ kill[l, s, t]

That is, the fluent *the unjust leader's killing the soldiers* holds at time t if and only if the proposition *the unjust leader killed the soldiers at time t* is true. Fluents are more like propositions, and hence unlike events they resist adjectival modification at a semantic level (see also Nikitina 2011).

The more consistently nominal MMC B would then represent reification of the underlying proposition into an event, or more strictly into an event-type. Event-types are not time-dependent in the same way as fluents, and are derived by Hamm and Lambalgen through existential closure of the time variable. That is, the event-type *the unjust leader's killing of the soldiers* would be represented as in (20a). Event-types can be arguments of the HAPPEN predicate, which converts them into event-tokens, as in (20b):

(20) a. ∃a[kill[l, s, a]
b. Happen (∃a[kill[l, s, a], t)

That is, the event-type *the unjust leader's killing of the soldiers* becomes an event-token when it happens at a particular time *t*. The HAPPEN predicate is not a truth predicate, like HOLD AT, and thus the direct relationship that we observed in (19b) between a fluent and the underlying proposition is absent.

The details of how fluents and event-types combine with general and temporally-specifying predicates, i.e. loose and narrow containers, are complex, and we refer the reader to Hamm and Lambalgen (2005) for a proposal. Essentially, the type of general predicates such as "surprised everyone" is argued to be such that they take a fluent and a time as their arguments, while temporally-specifying predicates like "occurred on 6th July" denote sets of event-tokens. Associated with each event-type there is also a corresponding canonical fluent, defined as the set of times at which the event-type happens. This essentially allows all event-types to be coerced into functioning as fluents, and hence as possible arguments of a general predicate.

5 Conclusion

In this paper we have argued for a new and alternative approach to mixed category nominalizations. The LFG head-sharing approach requires the postulation of a syntactic VP to house the apparently verbal components of these constructions, in particular accusative objects and adverbs. Instead, we propose that the purely selectional properties of such constructions might be handled at a functional rather than strictly categorial level. That is, a derived nominal will be allowed simply to take over in its entirety the argument structure of the verbal predicate from which it is derived, and thus license accusative objects **as a noun**. The potential for different kinds of modification, in particular modification by adjectives or adverbs, will ultimately depend on the semantic characteristics of the construction rather than its nominal category.

In particular, we have argued that the more mixed of the two Arabic masdar constructions. MMC A, is best treated as categorially nominal from top to bottom, just like the more consistently nominal MMC B. The prime indicator of MMC A is the presence of accusative object arguments. However, the postulation of a syntactic VP in MMC A to house these leads to issues

concerning constituency, and incorrectly predicts that adverb modification should be available solely to MMC A when it is in fact equally possible in MMC B.

One of the prime motivations for postulating the presence of a syntactic VP in the analysis of mixed category nominalizations appears to be an attempt to maintain a division between the complementization potential of verbal and nominal heads. However, the strictness of this division seems to be a meta-theoretical desire rather than necessarily an empirical reality. A second motivation is the observation that mixed category constructions generally manifest phrasal coherence (for the term see Malouf 2000): that is, in the case of mixed category nominalizations the apparently verbal elements of the construction are contiguous. We note that phrasal coherence indeed seems to hold of MMC A: the construction begins with the nominal CS, and this is followed first by the accusative object(s) and then by any further obliques or (non-adjectival) adjuncts. However, in the case of MMC A, this apparent coherence simply results from the addition of accusative objects to a general NP structure which is independently motivated, and indeed manifested in MMC B. The point at which the accusative objects are added, as sisters of the masdar, is a natural position for core arguments.

We leave it an open question whether this new approach to mixed category nominalizations can be applied to all such constructions. This seems not implausible. For example, in the Italian *infinito sostantivato* construction, the infinitive would be treated as categorically a noun which takes an accusative object, rather than as a verb. The potential for modification by compatible adverbs would follow, as in the masdar constructions examined here, from the semantic analysis, rather than from the presence of a syntactic VP. And the adverb would automatically have to be a post-head adjunct because adverbs are generally blocked from the pre-head modifier position in Italian NPs. More work would need to be done however to establish whether the Italian construction has a fluent interpretation or not, and how this relates to the potential for modification by compatible adjectives. Ultimately, the answer to the general question will depend on a detailed analysis of the semantic properties of the individual constructions, and the extent to which phrasal coherence properties can be argued to emerge as an epiphenomenon of more general and language-specific principles of ordering and constituency. Our goal in this paper has been merely to show the plausibility of the analysis for Arabic.

References

Al-Sharif, Ahmad M. 2014. *The syntax of negation in Arabic: an LFG perspective*. PhD dissertation, University of Essex.

Al-Sharifi, Budour and Louisa Sadler. 2009. The adjectival construct in Arabic. *Proceedings of the LFG09 Conference, Trinity College, Cambridge*, ed. by Miriam

Butt and Tracy Holloway King. On-line, CSLI Publications: http://www-csli.stanford.edu/publications/.

Bardeas, Suzanne M. 2009. T*he syntax of the Arabic DP*. PhD Dissertation, University of York.

Bresnan, Joan. 1997. Mixed categories as head-sharing constructions. *Proceedings of the LFG97 Conference, University of California, San Diego*, ed. by Miriam Butt and Tracy Holloway King. On-line, CSLI Publications: http://www-csli.stanford.edu/publications/.

Bresnan, Joan and John Mugane. 2006. Agentive nominalizations in Gĩkũyũ and the theory of mixed categories. In Miriam Butt, Mary Dalrymple, and Tracy Holloway King (eds) *Intelligent Linguistic Architectures: Variations on themes by Ronald M. Kaplan*. Stanford: CSLI Publications, 201-234.

Danon, Gabi. 2008. Definiteness spreading in the Hebrew construct state. *Lingua* 118, 872-906.

Falk, Yehuda N. 2001. Constituent Structure and Grammatical Functions in the Hebrew Action Nominal. *Proceedings of the LFG01 Conference, University of Hong Kong*, ed. by Miriam Butt and Tracy Holloway King. On-line, CSLI Publications: http://www-csli.stanford.edu/publications/.

Fassi-Fehri, Abdelkader. 1993. *Issues in the structure of Arabic clauses and words*. Dordrecht: Kluwer.

Feferman, Solomon. 1984: Toward Useful Type-Free Theories I. *The Journal of Symbolic Logic* 49, 75-111.

Hamm, Fritz and Michiel van Lambalgen 2005. Formal foundations for semantic theories of nominalisation, Berlin: ZAS papers

Haspelmath, Martin 1999. Explaining article-possessor complementarity: economic motivation in noun phrase syntax. *Language* 75.2: 227-43.

Kibort, Anna. 2014. Mapping out a construction inventory with (lexical) mapping theory. *Proceedings of the LFG14 Conference, University of Michigan, Ann Arbor*, ed. by Miriam Butt and Tracy Holloway King. On-line, CSLI Publications: http://www-csli.stanford.edu/publications/.

Malouf, Robert P. 2000. *Mixed categories in the hierarchical lexicon*. Stanford: CSLI Publications.

Nikitina, Tatiana V. 2008. *The mixing of syntactic properties and language change*. ProQuest, UMI Dissertation Publishing.

Payne, John, Rodney Huddleston and Geoffrey K. Pullum, 2010. The distribution and category status of adjectives and adverbs. *Word Structure* 3, No. 1, 31-81.

Ryding, Karin C. 2005. *A reference grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press.

Shanahan, Murray. 1997. *Solving the Frame Problem - A Mathematical Investigation of the Common Sense Law of Inertia*. Cambridge, MA., MIT Press.

Spencer, Andrew. 2005. Towards a typology of 'mixed categories'. *Morphology and the web of grammar. Essays in memory of Stephen G. Lapointe*, ed. by C. Orhan Orgun and Peter Sells. Stanford, CA: CSLI Publications. 95–138.

Vendler, Zeno. 1967. *Linguistics in Philosophy*. Cornell University Press, Ithaca, New York

MORPHOLOGY IN THE LFG ARCHITECTURE

Mary Dalrymple
University of Oxford

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King (Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

In line with the overall modular approach of LFG, we assume that the morphological component has its own internal structure and obeys universal and language-particular constraints on word formation that need not be shared by other levels of structure. Following Sadler and Spencer (2001), Kaplan and Butt (2002), Spencer (2006, 2013), and many others, we assume that the morphological component of the grammar associates a word form with a set of morphological features representing the structure and contribution of the word, often analyzed as identifying a slot in a paradigm. This view presupposes a *realizational* theory of morphology as proposed by, among others, Stump (2001, 2006, 2012); it is, however, compatible not only with explicitly paradigm-based models, but with any realizational theory which relates words to feature sets encoding their grammatical properties and structure, including finite state theories of morphology (Kaplan and Kay, 1994; Beesley and Karttunen, 2003). Here, we show how lexical entries for word forms are produced on the basis of input from a realizational morphological component.

1 The Morphology-Syntax Interface in LFG

In common with much LFG work, we assume that the morphological component of the grammar associates syntactic, semantic, and other information with word forms, producing lexical entries for word forms.

A note about terminology is in order. Morphologists often use the term 'lexical entry' to refer to information associated with a **lexeme** rather than a word form. Here, we use the term 'lexical entry' to refer to a **word form** (for example, the plural noun *dogs*) and its associated syntactic, semantic, and phonological information. We will use the term *lexemic entry* to refer to the pairing between a lexeme and the f-description encoding grammatical information that all word forms of the lexeme have in common (what Ackerman and Stump (2004) call the 'lexemicon').

1.1 The Lexical Entry: Grammatical Information Associated with Word Forms

We assume that the full lexical entry for the plural noun *dogs* contains at least the following information[1] (Dalrymple and Mycock 2011; Mycock and Lowe 2013; Dalrymple et al. 2015; see Bögel 2015 for a related proposal):

[†]Thanks to Miriam Butt, Ron Kaplan, John Lowe, Louise Mycock, and Andy Spencer for detailed comments on drafts, and to the audience at LFG15 for helpful discussion.

[1]Depending on theoretical assumptions, additional features such as PERS or CASE may also be required to be present.

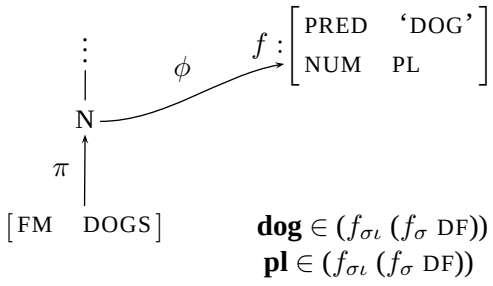
(1) Full lexical entry for *dogs*:

s-form	$(\bullet\ \text{FM}) =$ dogs
c-structure category	$\lambda(\pi(\bullet)) =$ N
f-description	$(\uparrow\ \text{PRED}) =$ 'dog'
	$(\uparrow\ \text{NUM}) =$ PL
	dog $\in$ $(\uparrow_{\sigma\iota}\ (\uparrow_{\sigma}\ \text{DF}))$
	pl $\in$ $(\uparrow_{\sigma\iota}\ (\uparrow_{\sigma}\ \text{DF}))$
p-form	/dɔgz/

We follow Dalrymple and Mycock (2011) and Mycock and Lowe (2013) in distinguishing two aspects of the string, the *s-string* and the *p-string*. The p-string for a word form is divided into prosodic units, each of which is related by rules of phonology and prosody to the p-form. In this lexical entry, the p-form is /dɔgz/. For more on prosodic structure and its representation, see Mycock and Lowe (2013), Dalrymple et al. (2015), and Bögel (2015).

The s-string is composed of s-string units. The s-string unit for a word form is represented in its lexical entry by the symbol $\bullet$, meaning the current s-string unit: its use is similar to the $*$ symbol standing for the current node of the phrase structure tree. Each s-string unit is an attribute-value structure containing the attribute FM whose value is a string representing the form of the word, as well as additional attributes and values which we will not discuss here. S-string units are related to terminal nodes of the c-structure tree via the projection function π, as shown in (2). In (1), the s-string unit contributed by the word form *dogs* is related to the c-structure node labelled with the category 'N', as specified in the second line of the entry: π is the function from s-string units to terminal nodes of the c-structure tree, and λ is the labelling function for c-structure nodes (Kaplan, 1995). The rest of the lexical entry contains f-structural information (specification of the PRED and NUM of *dogs*) and two meaning contributions, the meaning of the lexeme DOG and the semantic contribution of the plural morphology.

(2)



Our purpose in the following is to show how the s-form, p-form, c-structure category, and f-description for a word form are determined on the basis of the morphological structure of the word form, given a theory of morphological realization for the language. For simplicity, in the following exposition we will omit meaning

constructors and other nonsyntactic constraints in the f-description, working with a simplified lexical entry such as (3):

(3) Simplified lexical entry assumed here:

s-form	(● FM) =	dogs
c-structure category	$\lambda(\pi(\bullet))$ =	N
f-description	(↑ PRED) =	'dog'
	(↑ NUM) =	PL
p-form	/dɔgz/	

It is important to keep in mind that this simple f-description is a stand-in for the fully complete lexical entry, which encodes syntactic, semantic, information-structural, and other information by means of *templates* (Dalrymple et al., 2004) enabling the statement of generalizations across lexical entries, lexemic entries, and rule annotations within and across languages.

1.2 Lexical entries $\mathcal{L}$

Building on proposals by Kaplan and Butt (2002), we recast the different components of the lexical entry in (3) as a relation $\mathcal{L}$ involving an s-form, a p-form, a possibly complex c-structure category, and an f-description.

(4) $\mathcal{L}$<s-form, p-form, category, f-description>

The lexical entry given in (5) represents exactly the same information as in (3), but in a more convenient format for the definitions that we will provide.

(5) $\mathcal{L}$<dogs,
/dɔgz/,
N,
{(↑ PRED)='dog', (↑ NUM)=PL}>

1.3 Types of Morphological Features

We assume a realizational morphological component in which two types of morphological features are relevant. *Morphomic features* are relevant only for morphological realization, and play no role in other components of the grammar. The proper treatment of morphomic features and their role in morphological realization is an important issue in morphological theory, but since we do not depend on a specific theory of morphological realization, we will have nothing to say about morphomic features. A standard example of a morphomic feature is inflectional or declensional class.

M-features are any morphological features that have relevance for other components of the grammar: that is, any morphological features other than morphomic features (Sadler and Spencer, 2001; Spencer, 2006). We follow Sadler and Spencer

(2001) in prefixing morphological features with M-: for instance, writing the m-feature for morphologically encoded tense as M-TENSE, and the m-feature for morphologically encoded past tense as M-TENSE:PAST (Sadler and Spencer, 2001).

1.4 Lexemic Entries and the Lexemic Index

Spencer (2013) proposes that lexemic entries (which he calls 'lexical representations') have the following four components:

(6) Lexemic entries: Spencer (2013)

FORM:	the form of the root and any non-predictable stem forms
SYN:	syntactic information and requirements
SEM:	a representation of the meaning of the lexeme
LI:	a Lexemic Index, an arbitrary label identifying the lexeme

SYN and SEM constitute the f-description associated with the lexeme. We follow Spencer (2013) in assuming that each lexemic root is associated with a unique identifier, its Lexemic Index (similar to the LexID proposed by Stump 2001).

Building on Spencer (2013), we define a **lexemic entry** as a three-place relation LE involving (1) the form of the root and any non-predictable stem forms; (2) an f-description$_L$ that encodes syntactic, semantic, and other information associated with the lexeme, filling the role of Spencer's SYN and SEM; and (3) the Lexemic Index.

(7) General form of lexemic entry:

LE <root & idiosyncratic stem forms, f-description, Lexemic Index>

Lexemic entries for the lexemes with Lexemic Index DOG1 and CHILD1 are as follows:

(8) Lexemic entry for the lexeme DOG1:

LE <{ROOT:dog}, {(↑ PRED)='dog'}, DOG1>

(9) Lexemic entry for CHILD1:

LE <{ROOT: child; STEM1: children}, {(↑ PRED)='child'}, CHILD1>

The full f-description for a word form is obtained by combining the f-description$_L$ for the lexeme and the f-description$_M$ representing morphologically encoded grammatical information, as we will soon see.

1.5 The Realization Relation R

We assume that the morphological component specifies a morphological realization relation R, a set of four-place relations which we will call *m-entries*: R associates a Lexemic Index, an s-form, and a p-form with a set of m-features.

(10) General form of m-entry:

R <LexemicIndex, s-form, p-form, m-features>

For the word form *dogs*, we have the following m-entry:

(11) M-entry for the word form *dogs*:

R <DOG1, dogs, /dɔgz/, {M-CAT:NOUN, M-NUM:PL}>

We assume that the m-entries for each language are defined entirely by the morphological realization component R. The realization relation R for a language accounts for all aspects of the realization of word forms in the language, encompassing a theory of derivational and inflectional morphology, and encoding generalizations about affix ordering, stress placement, and other morphological patterns for the language. Our modular theory of the interface between the morphological component and the rest of the grammar makes no assumptions about the precise nature of R or the internal details of the morphological component; in the current context, R is simply a means of associating m-features with p-forms and s-forms relative to a lexemic root, and is compatible with any realizational theory of morphology.

1.6 The Description Function D

Finally, we require a means of interpreting the m-features for a word form as they are relevant to the rest of the grammar. We follow Kaplan and Butt (2002) in positing a description function D, which maps a set of m-features to the appropriate c-structure category and f-description$_M$, given a Lexemic Index (LI). D corresponds to what Andrews (2005) calls $\mathcal{F}$, and to what Sadler and Nordlinger (2004) call a "lexical transducer" relating m-features to grammatical specifications.

(12) General form of the description function D:

D <LI, m-features, category, f-description$_M$>

For the word form *dogs*, D maps the m-features {M-CAT:NOUN, M-NUM:PL} to the c-structure category N and the simplified f-description {(↑ NUM)=PL}:

(13) D <DOG1,
{M-CAT:NOUN, M-NUM:PL},
N,
{(↑ NUM)=PL}>

1.7 $\mathcal{L}$ Defined in Terms of D, LE, and R

We can now define $\mathcal{L}$ in terms of the set of lexemic entries LE, the morphological realization relation R, and the description function D which interprets the m-features to produce a c-structure category and f-description. $\mathcal{L}$ is the set of all lexical entries of the following form:

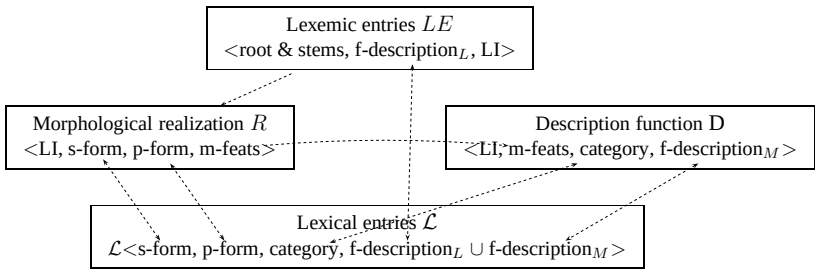
$<$s-form, p-form, category, f-description$_L \cup$ f-description$_M>$

which meet the conditions imposed by LE, R, and D:

(14) $\mathcal{L}$ = {$<$s-form, p-form, category, f-description$_L \cup$ f-description$_M>$:
LE $<$root (and idiosyncratic stem forms), f-description$_L$, LI$>$ $\wedge$
R $<$LI, s-form, p-form, m-features$>$ $\wedge$
D $<$LI, m-features, category, f-description$_M$ $>$}

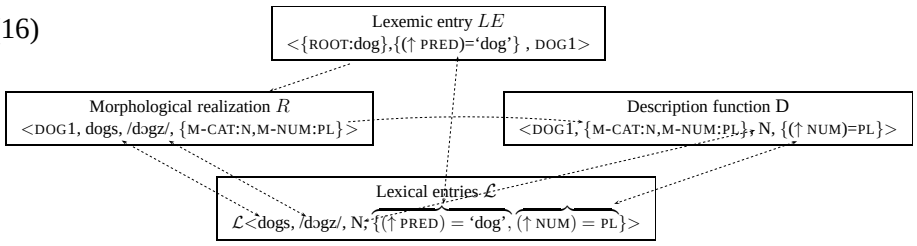
The diagram in (15) shows the flow of information in determining $\mathcal{L}$:

(15)



The diagram in (16) shows how the lexical entry for *dogs* is defined, given the other components:

(16)



Different features and structures are relevant and visible to different components of the grammar. Morphomic features are represented internal to R, and are not visible to $\mathcal{L}$. The Lexemic Index and m-features that are relevant for other components of the grammar are interpreted within $\mathcal{L}$, but do not appear in the lexical entries defined by $\mathcal{L}$ (Kaplan and Butt, 2002). This maintains a clean separation between morphomic features and other m-features, and between morphology and the other components of the grammar. In this way, the current proposal aligns itself with the Principle of Morphology-Free Syntax (Pullum and Zwicky, 1988; Zwicky, 1992), and contrasts with proposals that reject the Lexicalist Hypothesis, including Distributed Morphology (Embick and Noyer, 2007) and the Exo-Skeletal Model (Borer, 2013).

2 Morphological Features and Morphological Classes

In this section, we present some simple and informal examples of the D-mapping. The intention is to illustrate the range and types of D-mappings that may be required, given various alternative assumptions about the best way of treating a particular grammatical construction or the grammatical consequences of a particular

morphological alternation. In line with the overall architecture of LFG, our approach is modular in the sense that any particular theory of the realization relation R is generally compatible with a range of different possibilities for syntactic and semantic analysis. An example of this is presented in Section 2.1, where two alternative syntactic analyses of the English "affix hopping" pattern are considered in the context of the same theory of realization R for English. Our theory of the D-mapping as the interface between the morphological component and the rest of the grammar must be flexible enough to allow expression of alternative grammatical analyses on the basis of the same morphological realization relation R, and also to encompass alternative morphological assumptions about the nature of R, given a particular body of assumptions about the proper grammatical analysis of a syntactic construction.

We use $\stackrel{D}{\Rightarrow}$ informally for the D-mapping in this section. The full formal definition of D, to be provided in Section 3, covers all of the types of D-mappings to be examined in the rest of this section.

2.1 C-Structurally Relevant M-Features

English auxiliaries must appear in a particular linear sequence; this is the well-known "affix hopping" pattern (Chomsky, 1957):

(17) a. They are swimming.

b. They have swum.

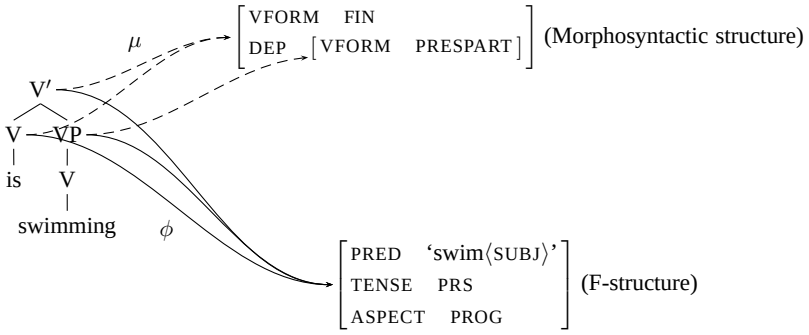
c. They have been swimming.

d. *They have are swum./*They have swimming.

Here we discuss two different LFG analyses of affix hopping. We do not take a position on which analysis is correct; rather, we use the two alternative analyses as illustrations of how the D-mapping works, and in particular to show how different definitions of the D-mapping from the same m-features give rise to different c-structure categories and f-descriptions depending on the syntactic analysis that is assumed.

Butt et al. (1996a,b) introduce a separate projection, *morphosyntactic structure*, reachable via the μ function from the c-structure. The role of morphosyntactic structure is to keep track of morphosyntactic dependencies such as affix hopping: on their analysis, embedding relations in morphosyntactic structure mirror embedding relations at c-structure, and the f-structure is monoclausal.

(18) *is swimming*, Butt et al. (1996a,b):



We assume that the present participle form *swimming* is associated via the realization relation R with the m-features M-CAT:VERB and M-VFORM:PRESPART:

(19) M-entry for the word form *swimming*:

R <SWIM1, swimming, /swɪmɪŋ/, {M-CAT:VERB, M-VFORM:PRESPART}>

On the analysis of Butt et al. (1996a,b), the m-feature M-VFORM:PRESPART corresponds to the feature VFORM with value PRESPART at morphosyntactic structure, and the feature ASPECT with value PROG at f-structure. For this analysis, the required D-mapping is given in (20):

(20) D mapping, Butt et al. (1996a,b):

M-VFORM:PRESPART $\stackrel{D}{\Rightarrow}$ {($\hat{*}_{\mu}$ VFORM)=PRESPART, (↑ ASPECT)=PROG}

On this analysis, then, a single m-feature (here, the m-feature M-VFORM:PRESPART) can map to an f-description consisting of more than one equation.

Frank and Zaenen (2002) present an alternative analysis of affix hopping in which morphosyntactic structure is projected from f-structure rather than from c-structure. Their analysis also appeals to complex c-structure categories such as V[fin,be], encoding fine-grained specifications over subtypes of standard categories like VP, V′, or V.[2] The parameters of a complex category are written in square brackets after the category label: a complex category like V[fin,be] is appropriate for a word form that is a verb (V) and can be categorized as finite (fin) and a form of the verb *be*. Parameter matching in c-structure rules allows for featural information to be passed through the c-structure, and for the c-structure position of words with particular parameters to be constrained. For example, if we assume that the first parameter of the V′ category can be instantiated to one of the three values 'fin', 'inf', and 'part', and furthermore that the V′'s parameter must match

[2]Miriam Butt [p.c.] notes that a similar complex-category-based analysis is implemented in the English Pargram grammar, and can be inspected through the INESS interface at `http://clarino.uib.no/iness/xle-web`.

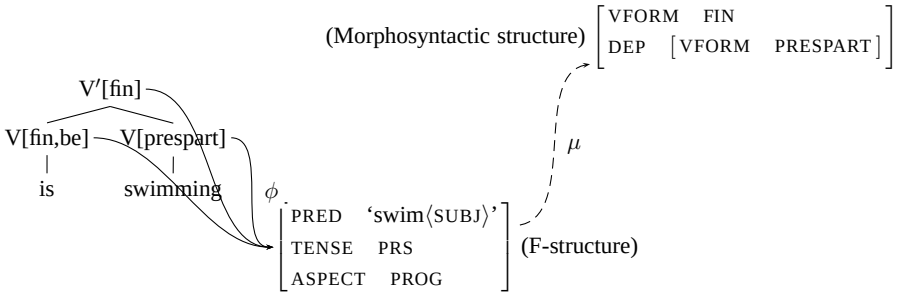
the first parameter of its V daughter, the rule in (21a) is a concise abbreviation of the three rules in (21b). All three of the rules in (21b) require a form of the verb *be* to be followed by a present participle form, and as such license a phrase such as *is swimming*.

(21) a. V′[_ftness] ⟶ V[_ftness,be] V[prespart]

b. V′[fin] ⟶ V[fin,be] V[prespart]
V′[inf] ⟶ V[inf,be] V[prespart]
V′[part] ⟶ V[part,be] V[prespart]

Under the assumptions of Frank and Zaenen (2002), the c-structure, f-structure, and morphosyntactic structure for *is swimming* is as in (22):

(22) *is swimming*, Frank and Zaenen (2002):



We assume that the finite form *is* has the following m-entry:

(23) M-entry for the word form *is*:

R <BE1, is, /ɪz/, {M-CAT:VERB, M-VTYPE:BE, M-VFORM:FIN}>

On Frank and Zaenen's analysis, the D-mapping maps from m-features to complex c-structure categories as well as determining the f-description. A combination of three m-features determines the complex c-structure category V[fin,be] for the word form *is*:

(24) D-mapping to complex c-structure category, Frank and Zaenen (2002):

{M-CAT:VERB, M-VTYPE:BE, M-VFORM:FIN} $\stackrel{D}{\Rightarrow}$ V[fin,be]

The M-VFORM:FIN feature is also involved in specifying the value of the VFORM feature at morphosyntactic structure:

(25) D-mapping to f-description, Frank and Zaenen (2002):

M-VFORM:FIN $\stackrel{D}{\Rightarrow}$ {($\uparrow_{\mu}$VFORM)=FIN}

Note that in the Frank and Zaenen (2002) analysis, more than one m-feature is involved in a single D-mapping relation, as in (24); furthermore, two different D-mapping relations may depend on the same m-feature, here, the M-VFORM feature, which appears in both (24) and (25). Our formal definition of D must therefore allow for these possibilities.

2.2 F-description$_M$ Dependent on Inflectional Class

It may sometimes be necessary for the D-mapping to depend on the simultaneous presence of more than one m-feature, and our formal definition of D must therefore be formulated to allow the mapping for each m-feature to be constrained by reference to the other m-features associated with a word form. For example, the past tense form of an English verb may be the same as its past participial form, or the two forms may differ. In what we will call Class 1, the past participle form is the same as the past tense form; in what we will call Class 2, there are two separate forms.

(26)

	ROOT	PAST TENSE	PAST PARTICIPLE
CLASS 1:	meet	met	
	talk	talked	
	make	made	
CLASS 2:	give	gave	given
	take	took	taken

One possible analysis of these patterns is to assume that inflectional class, represented as M-CLASS, is an m-feature which can be relevant in D-mapping. According to this analysis, the following m-entries are required:

(27) M-entries for English verb forms:

R <MEET1, met, /mɛt/, {M-CAT:VERB, M-ICLASS:1, M-TENSE:PAST}>

R <GIVE1, gave, /geɪv/, {M-CAT:VERB, M-ICLASS:2, M-TENSE:PAST}>

R <GIVE1, given, /gɪvən/, {M-CAT:VERB, M-ICLASS:2, M-FORM:PPART}>

Notice that on this analysis, there is only one m-entry for the Class 1 verb form *met*, associated with the m-feature M-TENSE:PAST. The D-mapping rules given in (28) produce the correct f-descriptions for these word forms: a disjunction specifying either past tense or the past participial form for Class 1 verbs like *meet*, and separate rules for past tense and past participial forms for Class 2 verbs like *give*.

(28) a. Past participle homophonous with past tense form in Class 1 verbs:

M-TENSE:PAST $\overset{D}{\Rightarrow}$ {{(↑ TENSE)=PAST | (↑ VFORM)=PPART}} in the presence of the m-feature M-ICLASS:1.

b. Separate past participle and past tense form in Class 2 verbs:

M-TENSE:PAST $\overset{D}{\Rightarrow}$ {(↑ TENSE)=PAST} in the presence of the m-feature M-ICLASS:2.

M-FORM:PPART $\overset{D}{\Rightarrow}$ {(↑ VFORM)=PPART}

If we adopt this analysis, M-ICLASS is not a morphomic feature relevant only for morphological realization, but is crucial in the definition of D in determining the full f-description for a verb form.

2.3 Lexical Exceptions: F-description$_M$ Dependent on Lexemic Index

We include the Lexemic Index as a component of the description function D in order to allow for the possibility that the interpretation of a set of m-features varies for different lexemes, and is thus dependent on the Lexemic Index. This is another kind of context-dependence that may be relevant in the definition of the D-mapping.

For example, Acquaviva (2008, page 19) proposes that the noun *measles* is exceptional in being morphologically plural (involving suffixation of plural *-s* to a base, and so carrying the m-feature M-NUM:PL) but syntactically singular, and so bearing the f-structure equation (↑ NUM)=SG:

(29) Measles is/*are a terrible disease.

Support for the view that *measles* is morphologically plural, consisting of the root *measle* followed by plural *-s*, is provided by attested examples of the uninflected form *measle* as the first member of a compound:

(30) a. New needle-free **measle vaccine** 'could save thousands of children's lives' (headline in *The Telegraph*, 17 Aug 2009)

b. ...it is reasonable because though we have never found a **measle germ** associated with **measle-symptoms** we have in cases with like symptoms found, not indeed **measle germs**, but things of the same sort... (Wisdom, 1968)

If we adopt this generalization, the D-mapping for a set of m-features may differ depending on the Lexemic Index LI. In the case at hand, we have one mapping involving the M-NUM feature for *measles*, and another mapping for all other nouns:

(31) *Measles* as a lexical exception:

M-NUM:PL $\overset{D}{\Rightarrow}$ {(↑ NUM)=SG} if LI = MEASLES1,
otherwise M-NUM:PL $\overset{D}{\Rightarrow}$ {(↑ NUM)=PL}.

This analytic possibility may or may not arise, depending on the particular theory of the morphology-syntax-semantics interface that is adopted. An alternative, more restrictive hypothesis is that D never takes the LI into account, and always relates a given set of m-features to the same f-description; this would require an alternative account of the difference between a syntactically singular noun like *measles* and a syntactically plural noun like *dogs*. If such an account is shown to be available and preferable for all word forms, we need not include the LI as a component of the description function D. Future work will show whether the more restrictive hypothesis is viable, or if the LI must be taken into account in at least some cases of the D-mapping.

2.4 Defaults in F-description$_M$

Some morphological theories assume defaults: that is, the absence of a feature is interpreted as indicating the presence of some grammatical property. Hence, the D-mapping must be formulated so as to allow for the possibility of introducing a particular f-description if no m-feature of a certain type is present.[3]

For example, we might propose that nouns are assumed to be syntactically singular if the plural m-feature does not appear, so that plural nouns are associated with the m-feature M-NUM:PL, and singular nouns lack a M-NUM feature.

(32) Singular number as a morphological default for nouns:

Introduce the f-description $\{(\uparrow \text{NUM})=\text{SG}\}$ if there is an m-feature M-CAT:N but no M-NUM m-feature in the m-description.

3 The Description Function D

3.1 Previous Definitions of D

In their analysis of the morphology-syntax interface, Kaplan and Butt (2002) assume that the D-mapping is defined very simply: the f-description corresponding to a set of m-features is constructed by examining one m-feature at a time, mapping it to a partial f-description independent of the presence or absence of other m-features. To illustrate their approach, they propose an analysis of the German noun *Kätzchen* which assumes that it is ambiguous, and that the correlation between number and case is captured by positing two different R-relations producing two different lexical entries, one with singular number and indeterminate NOM/DAT/ACC case, and the other with plural number and indeterminate NOM/GEN/DAT/ACC case:

(33) Kaplan and Butt (2002), *Kätzchen*:

a. $R <$KÄTZCHEN1, Kätzchen, /kɛts.çən/,
{M-ROOT:KATZE, M-CATEGORY:NOUN, M-DIMIN, M-GEND:NEUT,
M-NUM:SG, M-CASE:NOM/DAT/ACC}$>$

b. $R <$KÄTZCHEN1, Kätzchen, /kɛts.çən/,
{M-ROOT:KATZE, M-CATEGORY:NOUN, M-DIMIN, M-GEND:NEUT,
M-NUM:PL, M-CASE:NOM/GEN/DAT/ACC}$>$

Kaplan and Butt (2002) provide the following set of D-mappings from m-features to f-descriptions:

[3]Note that we neither advocate nor deplore the use of defaults in defining the D-mapping relation; whether or not defaults are needed in the D-mapping for a language depends on the characteristics of the morphological realization relation R and the syntactic rules and constraints for the language. Our aim is to formulate the D-mapping in a flexible enough way to allow for various alternative assumptions about the realization relation R and D-mapping for the language under analysis.

(34) Kaplan and Butt's (2002) description function D for each m-feature:

$$\begin{aligned}
D(\text{M-ROOT:KATZE}) &= \{(\uparrow \text{PRED}) = \text{`Katze'}\} \\
D(\text{M-CATEGORY:NOUN}) &= \{\text{N}, (\uparrow \text{NTYPE}) = \text{COUNT}\} \\
D(\text{M-DIMIN}) &= \{(\uparrow \text{DIMIN}) = +\} \\
D(\text{M-GEND:NEUT}) &= \{(\uparrow \text{GEND}) = \text{N}\} \\
D(\text{M-NUM:SG}) &= \{(\uparrow \text{NUM}) = \text{SG}\} \\
D(\text{M-NUM:PL}) &= \{(\uparrow \text{NUM}) = \text{PL}\} \\
D(\text{M-CASE:NOM/DAT/ACC}) &= \{(\uparrow \text{CASE}) \in \{\text{NOM,DAT,ACC}\}\} \\
D(\text{M-CASE:NOM/GEN/DAT/ACC}) &= \{(\uparrow \text{CASE}) \in \{\text{NOM,GEN,DAT,ACC}\}\}
\end{aligned}$$

On this simple view, there are no mappings from multiple m-features to a complex c-structure category (as we saw in 24), and the same m-feature cannot be involved in two different D-mapping relations (as we saw in 24 and 25). This view also does not allow for lexical exceptions taking into account the Lexemic Index (as we saw in 31), or for defaults in the D-mapping (as we saw in 32). Rather, the f-description corresponding to a set of m-features is just the union of all of the f-descriptions for each m-feature:

(35) Kaplan and Butt's (2002) description function D for a set of m-features:

$D(\{d_1, d_2, ..., d_n\}) = D(\{d_1\}) \cup D(\{d_2\}) \cup ... \; D(\{d_n\})$

This assumption is shared by Andrews (2005), who proposes a similar definition for his version of D, which he calls $\mathcal{F}$. This simple approach is adequate for many cases, but as we have seen, it is not adequate for all of the analytical possibilities that may arise.

3.2 Definition of D

We propose a definition of D that allows for the more complex cases examined in Section 2:

(36) D <LI, m-features, category, f-descr$_{default}$ $\cup$ f-descr$_{feat}$ > if and only if
D_{cat} <LI, m-features, category> and
$D_{default}$ <LI, m-features, f-descr$_{default}$ > and
D_{feats} <LI, m-features, f-descr$_{feats}$ >.

The subsidiary definitions D_{cat}, $D_{default}$, and D_{feats} are specified on a language-by-language basis, though there is likely to be a great deal of commonality in their definitions across languages; this is an important topic of research in the interface between morphology and other components of the grammar.

The (possibly complex) c-structure category for a word form is specified by D_{cat} on the basis of the L(exical) I(ndex) and the m-features. The f-description for the word form is determined by combining two subsidiary f-descriptions: D_{feats} contributes the f-description that is specified by the m-features associated with the word form, and $D_{default}$ contributes the default f-description that appears in the absence of certain marked m-features.

3.2.1 Defining the C-Structure Category

If complex categories are not assumed, the definition of D_{cat} is very simple, appealing to a straightforward specification of the c-structure category by the m-feature M-CAT:

(37) D_{cat} <LI, m-features, N> if and only if M-CAT:N $\in$ m-features.
D_{cat} <LI, m-features, V> if and only if M-CAT:V $\in$ m-features.
D_{cat} <LI, m-features, Adj> if and only if M-CAT:ADJ $\in$ m-features.
$\vdots$

If complex categories are assumed, more than one m-feature might be involved in the full specification of a complex category. Here is a representative example of a D_{cat} rule for the complex category V[fin,be] in the analysis of affix hopping shown in example (22):

(38) D_{cat} <LI, m-features, V[fin,be]>
if and only if $\{$M-CAT:VERB,M-VTYPE:BE,M-VFORM:FIN$\}$ $\subseteq$ m-features.

According to the definition of D in (36), D_{cat} is required to apply in order to determine the c-structure category of a word form. It does not do any "feature accounting", however; whether the definition of D_{cat} appeals to one m-feature or more than one, all of the m-features are passed on to $D_{default}$ to check for the application of rules involving privative m-features and defaults.

3.2.2 Privative M-features and Defaults

We assume that a set of default/privative D-mapping rules is defined for each language (including the possibility of no default mapping rules). Assuming that $D_{default}$ contains n default rules, the default f-description results from applying each of the n rules in turn:

(39) Default mappings $D_{default}$:

$D_{default}$ <LI, m-features, $d_1 \cup d_2 \cup ...d_n >$ if and only if
$D1_{default}$ <LI, m-features, $d_1 >$ and
$D2_{default}$ <LI, m-features, $d_2 >$ and
$\vdots$
$Dn_{default}$ <LI, m-features, $d_n >$.

The rule in (40) is a schematic rule illustrating the general form of default rules:

(40) Schematic default rule:

$D1_{default}$ <LI, m-features, $f_1 >$ if $m_1 \notin$ m-features (and possibly other conditions as well),

otherwise $D1_{default}$ <LI, m-features, $\emptyset >$.

For example, the following rule introduces a singular f-description for a noun that does not have a M-NUM:PL feature:

(41) Example: Default mapping to a singular f-description in the absence of a plural m-feature

$D1_{default}$ <LI, m-features, $\{(\uparrow \text{NUM})=\text{SG}\}$>
if M-CAT:N $\in$ m-features and M-NUM:PL $\notin$ m-features,
otherwise $D1_{default}$ <LI, m-features, $\emptyset$ >.

As with the D_{cat} rules, the $D_{default}$ rules are not involved in "feature accounting": once the default rules have applied to a set of m-features, that set is passed unchanged to the D_{feats} rule.

3.2.3 F-descriptions corresponding to m-features

Finally, the D_{feats} rule applies. This rule keeps track of features, and each m-feature must be accounted for by a D_{feats} rule. However, the D-mapping for an m-feature might correspond to an empty f-description: for example, a feature that specifies only c-structure category information might correspond to the empty f-description.

(42) Mapping m-features to f-descriptions:

D_{feats} <LI, $\{m_1, m_2, ...m_n\}$, $d_1 \cup d_2 \cup ...d_n >$ if and only if
D_{feats} <LI, m_1, $\{m_1, m_2, ...m_n\}$, $d_1 >$ and
D_{feats} <LI, m_2, $\{m_1, m_2, ...m_n\}$, $d_2 >$ and
$\vdots$
D_{feats} <LI, m_n, $\{m_1, m_2, ...m_n\}$, $d_n >$.

Simple D-mappings from m-features to f-descriptions An example of a simple mapping from the M-CASE feature to syntactic case is shown in (43):

(43) Example: Simple mapping from M-CASE to syntactic case

D_{feats} <LI, M-CASE:NOM, m-features, $\{(\uparrow \text{CASE})=\text{NOM}\}$>.
D_{feats} <LI, M-CASE:ACC, m-features, $\{(\uparrow \text{CASE})=\text{ACC}\}$>.
D_{feats} <LI, M-CASE:DAT, m-features, $\{(\uparrow \text{CASE})=\text{DAT}\}$>.
D_{feats} <LI, M-CASE:GEN, m-features, $\{(\uparrow \text{CASE})=\text{GEN}\}$>.

In formulating the D_{feats} rules for a language, it may be useful to appeal to a notational convention that allows for reuse of an m-feature value as the value of the corresponding f-structure feature, borrowing the underscore notation for the argument of a parametrized template (Dalrymple et al., 2004) to indicate that morphological case always matches syntactic case. Here, the value of the M-CASE feature is represented as _CASE with a preceding underscore, and is reused as the value of the f-structure feature CASE.

(44) Notational convention: General mapping from any M-CASE to the corresponding f-structure CASE specification, abbreviating the rules in (43)

D_{feats} <LI, M-CASE:_CASE, m-features, {(↑ CASE)=_CASE}>.

Context-sensitive D-mappings We have seen that a D-mapping rule for a particular m-feature may depend on the presence or absence of other m-features. For example, in Section 2.2 we saw that the f-description for a verb with m-feature M-VFORM:PAST can depend on the inflectional class of the verb. The following D_{feats} rule captures this dependency:

(45) Example: D-mapping dependent on the M-CLASS m-feature

D_{feats} <LI, M-VFORM:PAST, m-features, {{(↑ TENSE)=PAST | (↑ VFORM)=PPART}}>
if and only if M-CLASS:1 ∈ m-features.

D_{feats} <LI, M-VFORM:PAST, m-features, {(↑ TENSE)=PAST}}>
if and only if M-CLASS:2 ∈ m-features.

Vacuous D-mappings Since the D_{feats} mapping maps each individual m-feature to an f-description, we require a D_{feats} mapping rule for each m-feature, even those that do not correspond to an f-description. A schematic role for such inert m-features is the following:

(46) D-mapping to the empty f-description for an inert m-feature m_1

D_{feats} <LI, m_1, m-features, ∅ >.

For example, the M-CLASS feature may be important in controlling the mapping of other features (such as the M-VFORM feature, as shown in 45), but it may not itself correspond to any f-description; that is to say, it maps to the empty f-description ∅.

(47) Example: D-mapping to the empty f-description for the M-CLASS:1 and M-CLASS:2 m-features

D_{feats} <LI, M-CLASS:1, m-features, ∅ >.
D_{feats} <LI, M-CLASS:2, m-features, ∅ >.

For succinctness, we can introduce an additional notational convention: specification of an m-feature attribute like M-CLASS without specifying a value is interpreted as signifying all possible values for the M-CLASS feature.

(48) Example: D-mapping to the empty f-description for the M-CLASS m-feature with any value

D_{feats} <LI, M-CLASS, m-features, ∅ >.

D-mappings dependent on the LI Finally, the Lexemic Index is relevant for analyses involving lexically idiosyncratic D-mappings. Schematically, such analyses are of the following form:

(49) Schematic mapping for a lexical exception, dependent on the Lexemic Index:
D_{feats} <LI, m_1, m-features, f_1 > if LI = l_1,
otherwise D_{feats} <LI, m_1, m-features, f_2 >.

For example, we can treat the noun *measles* as a lexical exception, morphologically plural but syntactically singular:

(50) Example: Mapping to syntactically singular f-description for the morphologically plural noun *measles*, and to plural f-description for all other nouns
D_{feats} <LI, M-NUM:PL, m-features, {(↑ NUM)=SG}> if LI = MEASLES1,
otherwise D_{feats} <LI, M-NUM:PL, m-features, {(↑ NUM)=PL}>.

4 Conclusion

We have proposed a definition of $\mathcal{L}$ as the set of lexical entries for the word forms of a language. We rely on a set of lexemic entries LE and a morphological realizational component R which associates a set of m-features with a word form of a lexeme in the language. The description function D for the language maps from m-features to c-structure categories and f-descriptions. We hope that the proposals we have made will enable further exploration of the place of morphology in the architecture of LFG and the interface between morphology and the rest of the grammar.

References

Ackerman, Farrell and Stump, Gregory. 2004. Paradigms and Periphrastic Expression: A Study in Realization-Based Lexicalism. In Sadler and Spencer (2004), pages 111–157.

Acquaviva, Paolo. 2008. *Lexical Plurals: A Morphosemantic Approach*. Oxford: Oxford University Press.

Andrews, III, Avery D. 2005. F-Structural Spellout in LFG Morphology, unpublished manuscript, Australian National University.

Beesley, Kenneth R. and Karttunen, Lauri. 2003. *Finite-State Morphology*. Stanford: CSLI Publications.

Bögel, Tina. 2015. *The Syntax-Prosody Interface in Lexical Functional Grammar*. Ph. D.thesis, Universität Konstanz.

Borer, Hagit. 2013. *Structuring Sense, Volume III: Taking Form*. Oxford: Oxford University Press.

Butt, Miriam, Fortmann, Christian and Rohrer, Christian. 1996a. Syntactic Analyses for Parallel Grammars: Auxiliaries and Genitive NPs. In *Proceedings of the International Conference on Computational Linguistics (COLING96)*, pages 182–187, International Committee on Computational Linguistics, Copenhagen.

Butt, Miriam, Niño, María-Eugenia and Segond, Frédérique. 1996b. Multilingual Processing of Auxiliaries Within LFG. In Dafydd Gibbon (ed.), *KONVENS '96: Natural Language Processing and Speech Technology: Results of the 3rd KONVENS Conference*, pages 111–122, Mouton de Gruyter, reprinted in Sadler and Spencer (2004, 11–22).

Chomsky, Noam. 1957. *Syntactic Structures*. Mouton de Gruyter.

Dalrymple, Mary. 2001. *Lexical Functional Grammar*, volume 34 of *Syntax and Semantics*. New York: Academic Press.

Dalrymple, Mary, Kaplan, Ronald M. and King, Tracy Holloway. 2004. Linguistic Generalizations Over Descriptions. In Miriam Butt and Tracy Holloway King (eds.), *On-Line Proceedings of the LFG2004 Conference*, Stanford: CSLI Publications.

Dalrymple, Mary, Lowe, John and Mycock, Louise. 2015. Lexical Functional Grammar. Second edition of Dalrymple (2001), in preparation.

Dalrymple, Mary and Mycock, Louise. 2011. The Prosody-Syntax Interface. In Miriam Butt and Tracy Holloway King (eds.), *On-Line Proceedings of the LFG2011 Conference*, Stanford: CSLI Publications.

Embick, David and Noyer, Rolf. 2007. Distributed Morphology and the Syntax-Morphology Interface. In Gillian Ramchand and Charles Reiss (eds.), *The Oxford Handbook of Linguistic Interfaces*, pages 289–324, Oxford: Oxford University Press.

Frank, Anette and Zaenen, Annie. 2002. Tense in LFG: Syntax and Morphology. In Hans Kamp and Uwe Reyle (eds.), *How We Say WHEN It Happens: Contributions to the Theory of Temporal Reference in Natural Language*, pages 17–51, Tübingen: Niemeyer, reprinted in Sadler and Spencer (2004, 23–66).

Kaplan, Ronald M. 1995. The Formal Architecture of Lexical-Functional Grammar. In Mary Dalrymple, Ronald M. Kaplan, John T. Maxwell, III and Annie Zaenen (eds.), *Formal Issues in Lexical-Functional Grammar*, pages 7–27, Stanford: CSLI Publications, previously published in *Proceedings of ROCLING II*, ed. C.-R. Huang and K.-J. Chen, Tapei 1989, pp. 1–18, and in *Journal of Information Science and Engineering 5* (1989), pp. 305–322.

Kaplan, Ronald M. and Butt, Miriam. 2002. The Morphology-Syntax Interface in LFG, abstract in *On-line Proceedings of the LFG2002 Conference*, ed. Miriam Butt and Tracy Holloway King.

Kaplan, Ronald M. and Kay, Martin. 1994. Regular Models of Phonological Rule Systems. *Computational Linguistics* 20(3), 331–378.

Mycock, Louise and Lowe, John J. 2013. The Prosodic Marking of Discourse Functions. In Miriam Butt and Tracy Holloway King (eds.), *On-Line Proceedings of the LFG2013 Conference*, Stanford: CSLI Publications.

Pullum, Geoffrey K. and Zwicky, Arnold M. 1988. The Syntax-Phonology Interface. In Frederick J. Newmeyer (ed.), *Linguistics: The Cambridge Survey*, Cambridge, UK: Cambridge University Press.

Sadler, Louisa and Nordlinger, Rachel. 2004. Relating Morphology to Syntax. In Sadler and Spencer (2004), pages 159–186.

Sadler, Louisa and Spencer, Andrew. 2001. Syntax as an Exponent of Morphological Features. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology*, pages 71–96, Dordrecht: Kluwer.

Sadler, Louisa and Spencer, Andrew (eds.). 2004. *Projecting Morphology*. Stanford: CSLI Publications.

Spencer, Andrew. 2006. Syntactic vs. Morphological Case: Implications for Morphosyntax. In Leonid Kulikov, Andrej Malchukov and Peter de Swart (eds.), *Case, Valency and Transitivity*, pages 3–22, Amsterdam: John Benjamins.

Spencer, Andrew. 2013. *Lexical Relatedness: A Paradigm-Based Model*. Oxford: Oxford University Press.

Stump, Gregory T. 2001. *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge, UK: Cambridge University Press.

Stump, Gregory T. 2006. Heteroclisis and Paradigm Linkage. *Language* 82(2), 279–322.

Stump, Gregory T. 2012. The Formal and Functional Architecture of Inflectional Morphology. In Angela Ralli, Geert Booij, Sergio Scalise and Athanasios Karasimos (eds.), *On-line Proceedings of the Eighth Mediterranean Morphology Meeting (MMM8) Cagliari, Italy, 14–17 September 2011*, pages 254–268, Patras: University of Patras.

Wisdom, John. 1968. *Other Minds*. Berkeley: University of California Press.

Zwicky, Arnold M. 1992. Some Choices in the Theory of Morphology. In Robert D. Levine (ed.), *Formal Grammar: Theory and Implementation*, pages 327–371, Oxford: Oxford University Press.

EGYPTIAN ARABIC PERCEPTUAL REPORTS

Shaimaa ElSadek†[+] and Louisa Sadler†
[+]Alexandria University and †University of Essex

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King (Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

This paper offers what is to our knowledge the first discussion of the encoding of perceptual reports in Egyptian Cairene Arabic (henceforth ECA). We build on the analytic framework offered by Asudeh and Toivonen (2012) (henceforth AT), in particular their discussion of *individual* and *eventuality* type PSOURCEs, and also on the account of perceptual reports in different varieties of Arabic which is given in Camilleri et al. (2014). We show that a range of different syntactic constructions are used to encode perceptual reports in the *seem/appear* category. ECA does not make productive use of the verbs (such as *yabdū* 'seem') which are used for perceptual reports in Modern Standard Arabic, instead the range of forms used include the active participle *bāyen* and the noun *ʃakl*. This paper offers a description of the main means used for perceptual reports in ECA, illustrating a number of issues which arise for an eventual formal analysis. We consider how the distinction between individual and eventuality type PSOURCE introduced in Asudeh (2012) and Asudeh and Toivonen (2012) plays out across these constructions.

1 Background

1.1 Perceptual Reports

Asudeh (2004, 2012) and Asudeh and Toivonen (2012) develop an account of perceptual reports in LFG in which they distinguish two main classes of verbs in which an individual rather than some aspect of the eventuality serves as the directly perceptible source (PSOURCE) for the report: copy raising perceptual report verbs and perceptual resemblance verbs with thematic subjects.

They argue that in the English copy raising (CR) construction, illustrated in (1) and (2), the subject of the perceptual report verb is necessarily interpreted as the visible source of perceptual evidence: this requirement is absent from subject raising (SSR) and expletive (EXPL) uses of English subject raising verbs.

(1) Chris seemed **like he** enjoyed the marathon.

(2) John seems **like** Mary defeated **him**.

In English, the copy raising clause is mediated by *like/as though/as if*, and the subject of *seem* is associated with an obligatory pronominal copy in the embedding clause. Asudeh and Toivonen (2012) argue that the *individual* PSOURCE requirement (holding of the subject) means that the English copy raising construction with *seem* will be infelicitous if (some aspect of) this individual cannot be directly perceived.[1] Thus SSR and EXPL uses of perceptual report verbs such as English *seem* may be used in a wider set of circumstances: suppose that Kim, an incorrigible

[†]We thank participants at LFG2015, the editors, and the external reviewer for helpful comments and feedback.

[1]But see e.g. Landau (2011) for a different view. Landau argues that having an individual PSOURCE is not a necessary requirement for English CR structures. What is important for our pur-

user of pungent aftershave, has just left the room. Entering soon after, a colleague might remark *Kim seems to have been in here*, but *Kim seems like he's been in here* would be infelicitous because the source of the perception is just some aspect of the *eventuality*, in this case, a lingering characteristic odour. Asudeh and Toivonen (2012) argue that PSOURCE is not a *thematic argument* of the CR verb, but is an entailed participant *in the state that the verb denotes* , and therefore they introduce a type distinction. In non-expletive CR the PSOURCE is of type *individual* while in other perceptual reports the PSOURCE is of type *eventuality*. While in English, the individual PSOURCE is expressed as the subject of the CR predicate, with the Swedish verb *verka* 'seem', it may be expressed by means of a PP in the *verka* 'seem' clause.

1.2 Perceptual Resemblance Verbs

Perceptual resemblance verbs, such as English *look, sound, smell* are very similar, except that the individual PSOURCE is expressed as a thematic SUBJ, and hence a copy pronoun is not required in the embedded clause. As in CR, the embedding is introduced by *like/as though/as if* , and the subject (or an aspect associated with it) is interpreted as the individual PSOURCE.

(3) John looked/sounded/smelled **like** Bill had served asparagus.

(4) John looked/smelled **like** he'd been running.

1.3 English Copy Raising

The syntactic analysis of English copy raising presented in Asudeh and Toivonen (2012) maintains the syntax of standard SSR, treating the PSOURCE as a non thematic argument of *seem*. Although Asudeh and Toivonen (2012) do not provide argument structures in their PRED values (hence for them, the PRED value of *seem* is simply 'SEEM'), we assume the more complex semantic forms here for increased clarity and consistency with standard LFG accounts of subcategorisation.

(5) *seem* (↑ PRED) = 'SEEM< XCOMP> SUBJ'
(↑ XCOMP SUBJ) = (↑ SUBJ)

The intervening predicate *like* (equivalently, *as though* and *as if*) is treated as a predicative element which heads an XCOMP and itself subcategorises for a COMP argument:

(6) *like* (↑ PRED) = 'LIKE < SUBJ, COMP>'

poses here is less the details of any one analysis but rather the investigation of the distinction between *individual* and *eventuality* type PSOURCEs in constructions in ECA which have not received attention in the literature to date. We leave the development of a theoretical account of the patterns we see to subsequent work.

There is a standard anaphoric binding relation between the subject of the CR verb and the obligatory pronominal copy in the embedded clause. In semantic composition the pronominal resource is subsequently removed by a manager resource (as in resumption): since these details of the analysis are not important for our discussion here, we will not illustrate these aspects of their approach here. (7) illustrates the essence of the syntactic aspects of their analysis of English copy raising.

(7)
$$\begin{bmatrix} \text{PRED} & \text{`SEEM'} \\ \text{SUBJ} & a\begin{bmatrix}\text{PRED} & \text{`KIM'}\end{bmatrix} \\ \text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{`LIKE'} \\ \text{SUBJ} & \\ \text{COMP} & \begin{bmatrix} .. & p\begin{bmatrix}\text{PRED} & \text{`PRO'}\end{bmatrix}\end{bmatrix} \\ \text{PTYPE} & \text{CL-COMPAR}\end{bmatrix}\end{bmatrix}$$

1.4 Modern Standard Arabic

Taking this work as a starting point, Camilleri et al. (2014) looked at the expression of perceptual reports in Modern Standard Arabic (MSA) and Maltese. They show that MSA *yabdū* 'seem, appear' is a perceptual report verb which takes expletive subjects and complements introduced by the complementising particle *ʔanna* 'that' but does not permit SSR. In addition to the expletive construction, they argue that *yabdū* in fact exhibits a CR construction when the complement is introduced by the complementising particle *kaʔanna* 'as if' in place of *ʔanna*. The pronominal copy which is anaphorically related to the SUBJ of the perceptual report verb may occur in a wide range of nominal GF functions in the embedded clause. The following examples illustrate the expletive and the copy raising constructions respectively.[2]

(8) yabdū ʔanna l-ʔawlād-a qad ḥaḍar-ū
seem.IPFV.3SGM that DEF-boys-ACC PTL come.PV-3PLM
It seems that the boys have come. EXPL

(9) bad-at-i l-bint-u kaʔanna-hā
seem.PV-3SGF-INDIC DEF-girl-NOM as-if-3SGF.ACC
katab-at-i r-risālat-a
write.PV-3SGF-INDIC DEF-letter-ACC
The girl seemed as if she wrote the letter. CR Salih 1985: 138

In (9), the subject represents the PSOURCE, and the complement is introduced by *kaʔanna*. The pronominal copy (which is expressed by the verbal inflection and

[2]Note that both the complementising particles *ʔanna* and *kaʔanna* must be immediately followed by a nominal element, or take a pronominal inflection.

also as a pronominal inflection on the complementising particle *ka?anna*) functions as the subject of the embedded clause. Camilleri et al. (2014) suggest when the perceptual report verb *yabdū* occurs in a non-expletive construction with *ka?anna* the matrix subject is necessarily interpreted as the PSOURCE.

To summarise, this perceptual report verb in MSA shows an expletive subject construction and a copy raising pattern with an individual PSOURCE requirement, the latter occurring only in the presence of the complementising particle *ka?anna* introducing the embedded clause.[3]

1.5 ECA ʔiz-zaher

The closest equivalent to the MSA use of the predicate *yabdū* in an expletive construction in ECA is most likely the definite N *ʔiz-zaher* 'the apparent', followed by a sentence introduced by the complementiser *ʔin* 'that', a frequent and invariant usage. Like the use of *seem* with an expletive subject, this lacks any individual PSOURCE requirement.

(10) ʔiz-zāher ʔin mona gat
DEF-apparent that Mona come.PV.3SGF
It seems that Mona came.

In the following sections we discuss in detail two alternative means by which perceptual reports may be expressed in ECA. It is these two construction types which are the focus of the current paper.[4]

2 ECA bāyen

2.1 Expletive Pattern

First consider examples such as (11)-(12), in which we see the form *bāyen*. This is the active participle of the verb *bān* 'show/appear' in the default MSG form and is followed by a sentential complement introduced by the complementiser *ʔin* 'that'. The complement clause can in fact be introduced by either *ʔin* or *ka?in*. These complementising particles are amongst those which require the NP subject (if non-pronominal) to be immediately adjacent. If the subject is pronominal, the complementiser appears in inflecting form as shown in examples (11)-(13), and others. In the presence of the affixal material, the final consonant of the complementising particle geminates. In (11) the embedded sentence has an adjectival (participial) predicate while (12) contains a finite (bi-imperfective) verbal predicate.[5]

[3]Although in principle such examples could involve a thematic use of *yabdū* in the presence of *ka?anna*, Camilleri et al. (2014) reported that none of the examples used in the sources they consulted lacked a copy pronoun.

[4]Glosses used in this paper include AP 'active participle', BI.IPFV 'bi-imperfective, INDIC 'indicative', IPFV 'imperfective', PV 'perfective', PSP 'passive participle'.

[5]The bi-imperfective forms (glossed BI.IPFV) are finite forms, while the 'bare' imperfective (glossed IPFV) may be used in modal and what we take to be non-finite contexts.

(11) bāyen ʔinn-ik mabsūt-a
show.AP.SGM that-2SGF happy.PSP.SG-F
It seems that you (f) are happy.

(12) bāyen ʔinn-ik bitiʃrabi sagāyir
show.AP.SGM that-2SGF drink.BI.IPFV.3SGF cigarettes
It seems that you (f) smoke.

A perceptual report may be temporally located in the past by adding the tense auxiliary *kān* as in (13).

(13) kān bāyen ʔinn-ak mabsūt/bitiʃrab
be.PV.3SGM show.AP.SGM that-2SGM happy.PSP.SGM/drink.BI.IPFV.3SGF
sagāyir
cigarettes
It seemed that you're happy/It seemed that you smoke.

In these examples, the construction is not subject to an individual PSOURCE requirement — rather, the source of the perception could be any aspect of the event. Note also that the AP occurs in the invariant default MSG form, which is in contrast with other deverbal uses of the active participle, in which it inflects for gender and number of the subject. This observation supports the idea that the examples in (11)-(13) are expletive subject uses.

In order to collect a wider set of examples covering a range of possible structures for the expression of perceptual reports with *bāyen*, we compiled a corpus of ECA using the web crawling tools available within Sketchengine[6]. By directing the web crawl towards blogs, twitter feeds and other websites (such as magazine sites) known to originate with ECA speakers, a 1.5M word corpus of texts containing *bāyen* was constructed and manually checked for authenticity by the native speaker author. A corpus example showing the expletive use of *bāyen* is (14). Throughout this paper, the indicator SE-BYN shows that an example comes from our Sketchengine corpus.

(14) bāyen kont ba-ħeb ʔatfarrag ʕala ʔaflām
show.AP.SGM be.PV.1SG BI-love.IMPV.1SG watch.IMPV.1SG on movies
el-ʔakʃen
DEF-action
It seems that I used to love watching action movies. SE-BYN

Note that although (14) does not contain a complementiser, one can be added with no change in meaning:

[6]http://www.sketchengine.co.uk

(15) bāyen ʔinn-i kont ba-ħeb ʔatfarrag ʕala
show.AP.SGM that-I be.PV.1SG BI-love.IMPV.1SG watch.IMPV.1SG on
ʔaflām el-ʔakʃen
movies DEF-action
It seems that I used to love watching action movies.

In these examples, as with the previous constructed examples, the PSOURCE is not the individual, but rather some aspect of the event. A plausible context for (14) might be one in which the speaker is going through his old stuff and finds lots of action movies. Note that in (14) the auxiliary follows *bāyen*, and is in the embedded clause (in construction with *ba-ħeb*) 'used to love' and hence does not situate the 'seem' predication in the past. This is in contrast with example (13) where the auxiliary precedes *bāyen* and therefore the whole perceptual report is in the past tense (and hence the translation is 'it seemed').

2.2 Possible SSR Pattern

There are also cases in which the AP is not in the default SGM form, but in an agreeing form. This is the case in (16), where the AP is SGF, potentially indicating that *mona* is the subject of the AP. In such cases, it is again possible to use either *ʔin* or *kaʔin* to introduce the sentential complement.

(16) mona bayn-a kaʔinna-ha mabsūt-a
mona show.AP.SG-F that-3SGF happy.PSP.SG-F
Mona seems to be happy.

The perceptual report may be temporally located in the past by adding (a subject-inflected form of) the tense auxiliary *kān*, as in (17) and (18).

(17) kont bāyen ʔinn-ak mabsūt
be.PV.2SGM show.AP.SGM that-2SGM happy.PSP.SGM
You (m) seemed happy.

(18) kont-i bayn-a ʔinn-ik mabsūt-a
be.PV.2SG-F show.AP.SG-F that-2SGF happy.PSP.SG-F
You (f) seemed happy.

In such examples, it seems possible that *Mona* and *you* are subjects of the matrix (perceptual report) predicate (this is consistent with both the surface position of *Mona* in (16) and with the agreement inflection on the participle and the tense auxiliary). They are also most naturally interpreted as the PSOURCE, where the perceptual report is based on observation of some aspect of the individual. However, although these *may* be cases of SSR, we have found no parallel examples in our corpus data, and native speakers consulted judge them to be infrequent or unusual, although not ungrammatical.

The example in (19) is attested in our corpus: it shares with (16) the initial positioning of the NP. However this complementiser-less example clearly is not associated with an individual PSOURCE requirement, and corresponds very closely to the (constructed) example in (20).

(19) sāħeb el-ʃerka bāyen ʔeʃtara moʕzam ʔilli
owner DEF-company show.AP.SGM buy.PV.3SGM most that
ʔantag-u
produce.PV.3SGM-3SGM
It seems that the owner of the company bought most of what he produced. SE-BYN

This example seems rather to involve a misplacement (or very unusual placement) of the AP, where it's possible to have the AP initially, in the default SGM form followed by the complementiser, with no change in meaning, as in:

(20) bāyen ʔinn sāħeb el-ʃerka ʔeʃtara moʕzam ʔilli
show.AP.SGM that owner DEF-company buy.PV.3SGM most that
ʔantag-u
produce.PV.3SGM-3SGM
It seems that the owner of the company bought most of what he produced.

2.3 PP Pattern

In a further set of data, *bāyen* (in the default SGM non-agreeing form) co-occurs with a PP headed by *ʕala* 'on', which expresses the (individual) source of perception and usually corresponds to some (pronominal) function within the embedded CP.[7]

An example of this sort is shown in (21): note that the complementiser *kaʔin-* would be equally grammatical. (22) and (23) are corpus examples.

(21) kān bāyen ʕalē-ki ʔinn-ik mabsūt-a
be.PV.3SGM show.AP.SGM on-2SGF that-2SGF happy.PSP.SG-F
You seemed happy.

(22) bāyen ʕal-ēh ʔinn-u taʕbān giddan
show.AP.SGM on-3SGM that-3SGM tired.AP.SGM very
He seems to be very tired. SE-BYN

(23) kān bāyen ʕal-ēh ʔinn-uh masdūm
be.PV.3SGM show.AP.SGM on-3SGM that-3SGM shocked.PSP.SGM
He seemed to be shocked. SE-BYN

[7] Arabic is a language with inflecting prepositions. Pronominal objects of such prepositions are expressed inflectionally, while non-pronominal objects appear with the uninflected (default) form of the preposition. In all of examples (21)- (25) the preposition *ʕala* has a pronominal (inflectional) object while (26) illustrates a non-pronominal objects to *ʕala*.

Example (24) shows that although the AP is invariant (in default form), rather surprisingly the temporal auxiliary may optionally agree with the PSOURCE:

(24) konti bāyen ʕalē-ki ʔinn-ik mabsūt-a
be.PV.2SGF show.AP.SGM on-2SGF that-2SGF happy.PSP.SG-F
You seemed happy.

The object of *ʕala* 'on' within the matrix PP may correspond to a pronominal copy with a range of different grammatical functions in the embedded clause: in (25) it corresponds to the embedded object.

(25) bāyen ʕalē-ha ʔinn-ohom darab-ū-ha
show.AP.SGM on-3SGF that-3PL beat.PV-3PL-3SGF
She seems like they've beaten her.

It is also possible to have no copy pronoun in the complement clause, as in (26) below:

(26) bāyen ʕala el-modarreb ʔin el-Ahli keseb el-matʃ
show.AP.SGM on DEF-coach that DEF-Ahli win.PV.3SGM DEF-match
Lit: It seems on the coach that Ahli (football team) won the match. =
The coach seems like Ahli won the match.

2.4 Summary and Discussion

This section has illustrated a range of constructions involving the AP (active participle) *bāyen* to express perceptual reports in ECA. The AP *bāyen* can be used as the main predicate of perceptual reports in ECA. It exhibits three main behaviours and complementation patterns which might suggest different syntactic analyses. The first is the expletive variant which involves an invariant AP and a COMP introduced by *ʔin* or *kaʔin*. In this case, *bāyen* places no particular PSOURCE restriction, and hence the PSOURCE does not have to be the individual represented by the matrix subject, but can be any aspect of the event.

The second structure is the PP + CP variant, where the AP is followed by a PP which expresses the (visible) individual PSOURCE, irrespective of choice of the complementiser, which can be either *ʔin* or *kaʔin*. The AP must be in default form but a temporal auxiliary may agree with the nominal (PSOURCE) in the PP.

There is also the putative SSR variant, where the AP agrees in gender and number with the subject. This structure possibly has individual PSOURCE interpretations which may be associated with the choice of complementiser (*kaʔin* instead of *ʔin*) for the embedded complement. However this structure was much less frequent than the other two and appears to be of questionable grammaticality for some native speakers of ECA. We found no clear corpus examples representing this structure.

These data raise a number of interesting issues for further investigation. First, what is the nature of and significance of the use of the AP form?

A number of distinct uses of the AP can be distinguished in ECA and other vernaculars. Some are lexically specified as nouns where they have broken plural rather than sound plural forms[8]: for example, the plural AP *ʕummāl* 'workers' is related to the singular AP *ʕāmil* 'worker':

(27) ʔabelt el-ʕummāl
meet.PV.1SG DEF-workers
I met the workers. Mughazy 2004: 3

APs also occur in clearly adjectival function, occurring postnominal in the NP and showing concord in number, gender and definiteness with the head noun.

Beyond these cases, some APs, such as *lāzim* 'be necessary' and *gāyiz* 'be possible' are clearly modal. These occur with a single clausal argument and are always in default SGM form:

(28) lāzim ʔamʃi dilwaʔti
be necessary.AP leave.IMPV.1SG now
I must leave now. Mughazy 2004: 4

However although themselves always in default SGM form, such modal APs may be accompanied by default or agreeing forms of the temporal auxiliary.[9]

(29) kān lāzim tidʕi ʕali
be.PV.3SGM must.AP invite.IMPV.2SGM ali
You should have invited Ali (i.e. you did not). Azer 1980: 23

(30) kunt lāzim matballaġhūs el-χabar da
be.PV.2SGM must.AP inform.IMPV.2SGM.NEG DEF-news this
You shouldn't have told him this news. Azer 1980: 64

The use of an active participle as the main sentential predicate is completely standard in ECA, as in other contemporary vernaculars. The temporal/aspectual interpretation of the AP in such a deverbal usage depends on a number of factors which include the lexical aktionsart of the root verb and the presence or absence of temporal auxiliaries and adverbials: a very broad brush generalization is that APs from stative and motion verbs give rise to a range of meanings (including present progressive and various perfective meanings) while other (eventive) verbs give rise to perfective (typically, present perfect) meanings (see Mughazy (2004) for a very detailed discussion of how the range of interpretations can be accounted for). The following are typical examples of APs functioning as the main sentential predicate.

[8]The broken plural is formed by internal ablaut processes (as in English *mouse*, *mice*), while the sound plural involves regular suffixal morphology.

[9]To avoid confusion, these examples from Azer (1980) have been reglossed and re-transliterated to conform with the practice in this paper. We thank an anonymous reviewer for asking a question that caused us to provide these examples.

(31) mona lissa miχallaṣ-a el-wāgib
mona just finish.AP-SGF the-homework
Mona has just finished the homework. Mughazy 2004: 6

(32) ʕali sākin f-el-bēt da
Ali live.AP.SGM in-DEF-house this
Ali lives in this house. Mughazy 2004:6

As the example in (31) shows, a deverbal AP agrees in number and gender with its subject, which is different behaviour from what we have observed with the cases of *bāyen* in perceptual reports. The fact that the AP *bāyen* appears in default form in all the cases discussed above (with the exception of the potential SSR pattern illustrated in (16)-(18)) is therefore consistent with the idea that these perceptual report examples have expletive subjects (or no subject). Note further that the optional agreement of *kān* in the expletive subject (or no subject) modal examples such as (30) is also parallel to what we observe with *bāyen*.

Secondly, what is the status of the PP?

Asudeh and Toivonen (2012) show that the PSOURCE may be expressed with the Swedish verb *verka* 'seem' in one of two ways, either as a (raised) subject or as a PP (adjunct) with the expletive construction. The examples in (33) have the same meaning (where *Tom* and *han* are coreferential):

(33) a. Tom verkar som om han lagar mat
Tom seems as if he makes food
Tom seems as if he's cooking.

b. Det verkar på Tom som om han lagar mat
it seems on Tom as if he makes food
Tom seems as if he's cooking.

We have seen in the data discussed above that the ECA *ʕala*-PP also expresses the PSOURCE in an expletive construction.

Thirdly, what is the significance of agreement on the temporal Aux *kān?* Among the main possibilities to consider are that it represents a case of what we might call 'miscreant' or 'parasitic' agreement with a non-subject but prominent participant (the PSOURCE) or alternatively, that the PP (PSOURCE) is in fact the subject of the main predication.[10] In relation to the former possibility, we note that in other cases too *kān* can (optionally) show agreement with a prominent element which is not a clause-mate subject. In (34), for example, *kān* may optionally agree with the clause-mate object of the preposition (which is co-referential with the embedded subject).

[10]The term 'parasitic' agreement (or morphology) occurs in various places in the literature, including in discussion of some long-distance agreement effects (such as those in Hindi) in which agreement on an embedded infinitival is 'parasitic' on agreement expressed on the embedding predicate Boeckx (2008, 68). We also use the term here to denote agreement on a target which appears not to stand in a canonical agreement relationship with the controller.

(34) kont/kān maʕrūf ʕann-ak ʔinna-k kaddāb
be.PV.2SGM/PV.3SGM know.PSP.SGM about-2SGM that-2SGM liar
It was known about you that you're a liar/ you were known to be a liar.

But note that here too, the participle cannot show agreement:

(35) kānit maʕrūf/*maʕrūf-a ʕanna-ha ʔinna-ha
be.PV.3SGF know.PSP.SGM/know.PSP.SG-F about-3SGF that-3SGF
mabsūt-a
happy.PSP.SG-F
It was known about her that she was happy/She was known to be happy.

3 ECA ʃakl

3.1 ʃakl and Predicate

The following examples are both very natural and frequent as expressions of perceptual reports.

(36) ʃakl el-walad taʕbān
form DEF-boy tired.AP.SGM
The boy seems tired.

(37) ʃakl-ohom mestaneyīn ħāga mohemma
form-3PL wait.AP.PL thing important
They seem to be waiting for an important thing =
It seems they're waiting for an important thing.

They involve the noun *ʃakl* (MSG) which means 'form, shape' heading a construct state or *iḍāfa* construction (CSC) (in (36) above, the NP *ʃakl el-walad*), together with a following predicative phrase. The construct state is a common structure in ECA, and all other varieties of Arabic, in which a head noun, which cannot be inflected for definiteness, is immediately followed by a definite or indefinite nominal dependent, which is inflectionally expressed when pronominal. The CSC is the most common means of expressing possession (as in (38)) and a range of other associated relations such as part whole relations between head and dependent NP.

(38) a. kitāb mona
book.SGM mona
Mona' s book

b. kitab-ha
book.SGM-3SGF
her book

The status of the CSC as definite or indefinite depends on the definiteness of the (most deeply embedded) dependent nominal. There is a tight syntactic link between the head noun and the nominal dependent, such that adjectival modifiers (or complements of the head noun) may not intervene between them, but occur after the nominal dependent: adjectives agree with the noun (head or dependent) which they modify (note that in (39) the adjective is definite because the CSC as a whole is definite since the dependent nominal is definite):

(39) kitāb mona el-ʔadīm
book.SGM mona DEF-old.SGM
Mona's old book

(40) kitāb el-bint el-gamīla
book.SGM DEF-girl.SGF DEF-beautiful.SGF
the beautiful girl's book

Impressionistically, this structure with *ʃakl* is very frequently the usage of choice for the expression of perceptual reports. Nevertheless, the majority of instances of *ʃakl* which came up in our corpus searches were cases which the noun was being used simply in its standard meaning of 'form' or 'shape'. Having discounted these as irrelevant to our present concerns, it emerged that the vast majority (and perhaps all) of the corpus examples remaining contained an attached pronoun rather than a NP dependent, as in (37) or (41) (the latter contains an initial NP doubling the attached pronoun).

(41) Morsi ʃakl-u rigiʕ
Morsi form-3SGM come.back.PV.3SGM
It seems that Morsi came back.

It is rather clear that in (41) the PSOURCE is not of type *individual* as the Morsi in question is the ousted president of Egypt. The PSOURCE here is some aspect of the eventuality - something which is happening in the country that gives rise to this observation.

There are a number of interesting aspects to this perceptual report construction, which has not been discussed in any previous literature we are aware of.

The first issue concerns the nature of the (predicative or propositional) element which co-occurs with the CSC *ʃakl* NP. In examples (36) and (37) we see adjectives and verbal participles, which may head predicative complements, but example (41) exhibits a finite (perfective) verb (and hence potentially a full finite IP). Finite verbal complements are only possible provided the subject of the finite verb is co-referential with the dependent NP. This is illustrated by the contrast between (42) and (43).

(42) ʃakl el-welād etdarabo
form DEF-boys beat.PV.PASS.3PL
The boys seem to have been beaten.

(43) *ʃakl el-welād darabet-hom
form DEF-boys beat.PV.3SGF-3PL
The boys seem as if she's beaten them.

Note that in an example such as (42), the finite verb *etdarabo* beat.PV.PASS.3PL necessarily agrees with the dependent NP (within the CSC, that is the 'possessor' or 'whole'). This is a behaviour which is not found in other CSC, where it is the head noun (possessed element) which controls NP external predicate agreement. Similarly, it is the dependent NP which dictates the plural number on the deverbal AP in (37). In the adjectival case in (44), however, the dependent pronoun (or NP) only optionally controls agreement on the (embedded) predicate.

(44) ʃakl-ik mabsūt-a/mabsūt
form-2SGF happy.PSP.SG-F/PSP.MSG
You (f) seem happy.

As shown in (44) above, the dependent NP may control agreement on the predicate, a behaviour which is not found in other CSC, where it is the head noun (possessed element) which controls predicate agreement. Consider now (45)- (48): these examples show that when the dependent NP controls agreement on the predicate, it may also optionally control agreement on a temporal auxiliary in initial position.

(45) kān ʃakl-ik mabsūta
be.PV.3SGM form-3SGF happy.PSP.SG-F
You seemed happy.

(46) konti ʃakl-ik mabsūta
be.PV.3SGF form-2SGF happy.PSP.SGF
You seemed happy.

(47) kān ʃakl-ik biṫħibī-h
be.PV.3SGM form-2SGF love.BI.IPFV.2SGF-3SGM
Your form seemed to love him= You seemed to love him.

(48) konti ʃakl-ik biṫħibī-h
be.PV.2SGF form-2SGF love.BI.IPFV.2SGF-3SGM
You seemed to love him.

If the temporal auxiliary appears after the CSC NP (between the subject and the predicate) agreement with the dependent (annexed) NP within the CSC becomes obligatory.[11]

[11] These examples correspond to normal word order in ECA, which is SVO.

(49) *ʃakl-ik kān mabsūt-a
form-2SGF be.PV.3SGM happy.PSP.SG-F
Your form seemed happy/ You seemed happy.

(50) ʃakl-ik konti mabsūt-a
form-2SGF be.PV.2SGF happy.PSP.SG-F
You seem as if you were happy.

(51) ʃakl el-welād kānu biyitderbo
form the-boys be.PV.3PL beat.BI.IPFV.PASS.3PL
The boys seem to have been (being) beaten.

(52) ʃakl el-welād kānu tūl el-sana biyitderbo
form the-boys be.PV.3PL all the-year beat.BI.IPFV.PASS.3PL
The boys seem to have been (being) beaten all year.

Although the dependent NP (within the CSC headed by *ʃakl*) can (and sometimes must) control predicate agreement, it is not necessarily interpreted as the PSOURCE. In terms of the distinction drawn by Asudeh and Toivonen (2012) the PSOURCE is not of type *individual* in these cases, but just some aspect of the wider *eventuality*. Consider the following corpus example in this connection.

(53) ʃakl-ena keda ħa-nedxol ʕala ʕamaleyyet ʔalb maftūħ
form-1PL this FUT-enter.1PL on operation heart open.PSP.SGM
It seems that we'll get into an open heart operation. SE-SKL

Here the first interpretation that comes to mind is **not** the one in which 'we' refers to the group of doctors commenting on what they are actually doing. A plausible context is one in which the speaker is observing the operation or the patient or drawing a conclusion from the fact that the doctors are taking so long discussing options: the 'we' in such scenarios refers rather to people related to the patient in some way, perhaps those who might be affected by him having this operation. The same is true of (54), in which the annexed pronoun (dependent NP within the CSC) and the embedded subject are co-referential, referring to those who will make the arrest, but they do not have to be visible to the speaker (the speaker may, for example, be reporting on a rumour which is currently circulating). Hence, the PSOURCE is not required to be of type *individual* for (54).

(54) ʃakl-ohom ħayaʕtaqilū-h
form-3PL arrest.FUT.3PL-3SGM
It seems that they're going to arrest him.
They seem to be going to arrest him. SE-SHL

Another rather interesting structure is shown in (55), where the embedded verb is an OBJ-PSYCH verb (hence literally 'They were lacking to you') and the 'possessor' in *ʃakl-uhum* corresponds to the theme argument **not** the experiencer argument, while it is plausibly the appearance/behaviour of the experiencer which gives rise to the reported perception.

(55) ʃakl-uhum waħaʃū-k
form-POSS.3PL lack.PV.3PL-2SGM
It seems like you missed them! SE-SHL

The key points shown in this section are (i) the dependent NP/pronoun in the *ʃakl* CSC may (and sometimes, must) control agreement of a temporal auxiliary and of the predicate but (ii) is not obligatorily interpreted as the PSOURCE. In the following section, we consider a structure in which the embedded predicate is a CP introduced by a complementising particle.

3.2 kaʔin Complements

We have now seen examples in which the CSC headed by *ʃakl* is followed by an adjectival or deverbal participial predicative phrase, and cases in which it is followed by a finite predication. It may also occur with a sentential (CP) complement, which must be introduced by the complementising particle *kaʔin*. Notice that in this case, there is no requirement of co-reference between the dependent nominal inside the CSC and the subject of the (embedded) predication.

(56) ʃakl-aha kaʔinn-uhum deħku ʕalē-ha
form-3SGF as.if-3PL laugh.PV.3PL on-3SGF
She seems as if they've fooled her.

(57) ʃakl el-welād kaʔenn-aha darabet-hom
form DEF-boys as.if-3SGF beat.PV.3SGF-3PL
The boys seem as if she's beaten them.

(58) ʃakl-ak kaʔinn-ak mabsūt
form-2SGM as.if-2SGM happy.PSP.SGM
You seem as if you're happy.

There is an important and clear difference between these examples, involving a CP with the complementising particle *kaʔin*, and the previous examples with *ʃakl*, which do not involve a complementiser: (56)-(58) and other examples with this pattern receive an interpretation in which some aspect of the individual denoted by the dependent NP in the CSC is the source of the perceptual report. That is, these examples involve an *individual* PSOURCE.

The contrast between the following examples supports the observation that it is essentially the presence of the complementising particle *kaʔin* which introduces the

individual PSOURCE requirement. (59) requires direct observation of the patient, while in (60) it can be any aspect of the eventuality: for example, the reaction of her mother, or the lack of medicine in her room.

(59) kān ʃakl-aha kaʔenn-aha ʔetħassinit
be.PV.3SGM form-3SGF as.if-3SGF got.better.PV.3SGF
She seemed as if she got better. SE-SHL

(60) ʃakl-aha ʔetħassinit
form-3SGF got.better.PV.3SGF
It seems she's got better.

An association between *kaʔin* and some form of direct perception is found in other examples in which a NP occurs with a *kaʔin*-CP. Thus consider (61), with a CSC as subject and (62), with a simple NP subject. Each of these is interpreted as reporting a direct observation of the office (as indicated by the English free translation), and in fact substitution of the 'standard' complementising particle *ʔin* is impossible. [12]

(61) maktab-ak kaʔinn-u lessa mitratteb
office-2SGM as.if-3SGM just tidy.PSP.SGM
Your office looks as if its just been tidied up.

(62) el-maktab kaʔinn-u lessa mitratteb
DEF-office as.if-3SGM just tidy.PSP.SGM
The office looks as if its just been tidied up.

(63) *el-maktab ʔinn-u lessa mitratteb
DEF-office that-3SGM just tidy.PSP.SGM
The office that its just been tidied up.

Although we do not provide any details of an analysis here, we can view the relation between the CSC and the following *kaʔin* CP as mediated by an (unexpressed) copula predicate. It is no surprise then that we get a temporal auxiliary if the perception reported is temporally located in the past, as illustrated in (64). As with the other examples in this subsection, in (64) some aspect of the directly perceived corpse is the source of the perceptual report.

(64) kān ʃakl-o kaʔenn-o mayyet men fatra tawīla
be.PV.3SGM form-3SGM as.if-3SGM dead from time long
He seemed as if he's been dead for a long time.

[12] There is, of course, a crucial additional factor distinguishing (61), for example, from examples such as (58) and (57): in the latter examples there is an inalienable relation between the 'part' (head N) and the 'whole' (dependent NP). (61) would be totally infelicitous uttered when 'you' but not 'the office' was visible.

Agreement on the auxiliary is optional, where it can be controlled either by *ʃakl* or the annexed (dependent) nominal, with no apparent difference in meaning.

(65) kān ʃakl-aha kaʔenn-aha ʔetħassinit
be.PV.3SGM form-3SGF as.if-3SGF got.better.PV.3SGF
She seemed as if she got better. SE-SHL

(66) ʃakl el-welād kān kaʔenn-ohom etdarabo
form DEF-boys be.PV.3SGM as.if-3PL beat.PV.PASS.3PL
The boys seemed as if they were beaten.

(67) ʃakl el-welād kānu kaʔenn-ohom etdarabo
form DEF-boys be.PV.3PL as.if-3PL beat.PV.PASS.3PL
The boys seemed as if they were beaten.

3.3 Summary and Discussion

We have shown in this section that a natural and frequent means to express perceptual reports in ECA involves the use of a CSC headed by the N *ʃakl* 'form' and either a predicative phrase, finite IP or sentential complement introduced by the complementising particle *kaʔin*. The construct headed by *ʃakl* has unusual agreement properties, where the complement sometimes agrees with the dependent N (rather than the head N). However the pattern of agreement which we see does not correlate with the distinction between *individual* and *eventuality* types of PSOURCE: agreement patterns are basically independent of whether the dependent (annexed) nominal serves as PSOURCE. We have seen two further clear patterns in the data. First, a bare finite complement may occur only when the subject is co-referential with the dependent N within the CSC. Second, an individual PSOURCE requirement emerges only when *ʃakl* CS combines with a COMP, which must be introduced by *kaʔin*. The central role of this complementising particle in the construction is consistent with, and similar to, the role which Camilleri et al. (2014) argue is played by the cognate particle in MSA copy raising with the verb *yabdū* 'seem'.

4 Summary

This paper has presented a preliminary investigation of the expression of perceptual reports in ECA, based on a substantial corpus of data collected using the webcrawling corpus collection tools available in SketchEngine. We have discussed two constructions which are used to express perceptual reports in ECA. The first, less frequent, uses the AP *bāyen*. We have suggested that *bāyen* occurs principally in two constructions. The first is an expletive construction, and here there is no requirement that the PSOURCE be a (visible) individual. In the second construction, the PSOURCE is expressed in a PP, which we take to be in the matrix clause. This is reminiscent of the expression of an individual PSOURCE in a PP in Swedish,

discussed by Asudeh and Toivonen (2012). This construction has some interesting agreement properties, which we are tempted to view as cases of parasitic agreement. That is, predicate agreement is not always a reliable subject diagnostic. The second, more frequent way of expressing perceptual reports involved the use of the noun *ʃakl*, heading a (nominal) construct state construction and co-occurring with a predicative phrase, IP or a CP. The choice of complementising particle for CP in this construction is limited to *kaʔin*, and we argue that an individual P-SOURCE interpretation arises obligatorily, as in English CR, in such cases. Where the *ʃakl* CSC occurs with an IP/finite verbal complement, then the dependent nominal within the CSC and the subject of the finite verb are co-referential. With the *ʃakl* constructions we have also seen some interesting agreement facts, notably cases in which the dependent nominal within the NP CSC may control predicate agreement at clausal level.

References

Asudeh, Ash. 2004. *Resumption as Resource Management*. Ph.D. thesis, Stanford University.

Asudeh, Ash. 2012. *The Logic of Pronominal Resumption*. Oxford: Oxford University Press.

Asudeh, Ash and Ida Toivonen. 2012. Copy raising and Perception. *Natural Language and Linguistic Theory* 30/2:321–380.

Azer, Hany Amin. 1980. *The expression of Modality in Egyptian Colloquial Arabic: its syntax and semantics*. Ph.D. thesis, SOAS, University of London.

Boeckx, Cedric. 2008. *Aspects of the Syntax of Agreement*. London: Routledge.

Camilleri, Maris, Shaimaa ElSadek, and Louisa Sadler. 2014. Perceptual Reports in (varieties of) Arabic. In M. Butt and T. H. King, eds., *Proceedings of the LFG14 Conference*, pages 179–199. CSLI Publications: http://www-csli.stanford.edu/publications, Stanford, CA.

Landau, Idan. 2011. Predication vs aboutness in copy raising. *Natural Language and Linguistic Theory* 29(3):779–813.

Mughazy, Mustafa. 2004. *Subatomic semantics and the Active Participle in Egyptian Arabic*. Ph.D. thesis, University of illinois at Urbana-Champaign, Urbana-Champaign.

Salih, Mahmud. 1985. *Aspects of Clause Structure in Standard Arabic: A Study in Relational Grammar*. Ph.D. thesis, SUNY at Buffalo.

Soltan, Usama. 2007. *On Formal Feature Licensing in Minimalism: Aspects of Standard Arabic Morphosyntax*. Ph.D. thesis, University of Maryland.

NUMBER IN MERYAM MIR

Stephen Jones
University of Oxford

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King (Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

Meryam Mir is an Eastern Trans-Fly Papuan language spoken in the Torres Straits. It has four number categories which are exhibited through complex morphosyntax, with marking principally on verbs. Accounts of the language to date (Piper, 2013; Ray, 1907) have considered only argument number. Reanalysis of a fragment of the language demonstrates that number phenomena can be better explained by including verbal number (Durie, 1986) in the account. The reanalysis also suggests that Arka's assumption of a single feature set for argument and verbal number in Marori (Arka, 2012) does not hold for Meryam Mir. For argument number, it demonstrates that the Meryam Mir number categories and morphological alternations can be generated by a feature set including the novel feature [±GROUP]. The proposed feature set is compatible with those proposed by Sadler (2011) for Hopi as well as Arka for Marori (with some adaptation). For verbal number expressed through verb stem alternations, the feature [±BOUNDED] is proposed and tested. Further work is required to account for other verbal number phenomena and the small amount of data currently available leaves some questions unresolved.

1 Introduction

Meryam Mir is an endangered Papuan language with around 20 native speakers, one of four recorded from the Eastern Trans-Fly family, spoken on Mer Island and neighbouring islands in the eastern Torres Strait. It has basic SOV[1] word order, with split ergative case marking. There are four number categories, singular, dual, paucal, and plural, expressed predominantly through multiple agreement on verbs. Originally described by Ray (1907), a fuller account of the language has been given by Piper (2013).

This paper considers a fragment of the grammar: number marking by inflectional affixes and verb stem alternations. It establishes the presence of verbal number (plurality of events as opposed to plurality of event participants) as a distinct category in Meryam Mir with features that are also distinct from those underlying argument number. It proposes a novel argument feature set for the language, consistent with other LFG accounts of number features (Sadler, 2011; Arka, 2012) and also proposes one possible feature involved in the expression of verbal number.

[†]I thank Mary Dalrymple, who supervised the Masters thesis from which this paper is summarised, and Nick Piper, who collected and published the bulk of data on which the thesis is based. The paper has been greatly improved by the comments of the editors and an anonymous reviewer.

[1]The following abbreviations are used. A: Agent; ABS: Absolutive; AUG: Augmented; DU: Dual; FUT: Future; GenPL: Generic plural; HUM: Human; IPFV: Imperfective; LFG: Lexical Functional Grammar; LgPL: Large plural; LimPL: Limited plural; NP: Noun phrase; NPL: Nonplural; NSG: Nonsingular; NUM: Number; O: Object; OBJ: Object; PC: Paucal; PL: Plural; PRED: Predicate; PrN: Pronoun; PRS: Present; S: Subject; SG: Singular; SOV: subject-object-verb; SUBJ: Subject; *V*: Unspecified vowel; VNUM: Verbal number; VPL: Verbally plural; VSG: Verbally singular;

2 Expression of number in Meryam Mir

Meryam Mir has four number categories: singular (SG), dual (DU), paucal (PC) and plural(PL).[2] Paucal number is available only to human and high animate arguments. Nouns do not generally mark number, although optional number suffixes *-ey* and *-ba* mark noun phrases with DU number, and nouns that are members of groups respectively. Pronouns obligatorily mark a singular-nonsingular (NSG = DU/PC/PL) distinction.

Number marking on verbs includes inflection at prefix, infix and suffix sites. Suffixes usually mark subject number.[3] Prefixes and infixes mark the number of the absolutive argument (intransitive S, transitive O). Where the absolutive argument is O or inanimate/low animacy S, the prefix/infix number distinction is SG-NSG. However, where the absolutive argument is high animacy S, the unusual distinction SG/PL-DU/PC is observed.

There is also a set of verbs with stem alternations that are frequently, though not exclusively, associated with the number of the verb's absolutive argument. The alternating forms may be morphologically related[4] or may be suppletive pairs. The relevant verbs include examples from both morphological classes of verbs identified by Piper, Group I (atelic verbs, all intransitive) and Group II (telic verbs, either transitive or intransitive).

Turning first to the variants of number distinction: some verbs have a single root form for all number categories. These verbs come from Group I or Group II and do not seem to form a coherent semantic category.

Table 1: Verb stems with no alternation for number

Group I		Group II	
ikasir	'be going along'	*detager*	'tell'
emer	'nonhuman be sitting'	*ero*	'eat'
og	'climb'	*iri*	'drink'

A second type of distinction is between NPL and PL argument number. All verbs showing this distinction are intransitive with many of from Group I, though there are also Group II verbs that follow this pattern.[5]

A third type of distinction is between SG or DU argument number on the one hand, and PC or PL argument number on the other. The verbs showing this distinc-

[2] A fuller description of number marking is given in Piper (2013).

[3] Piper describes the operation of a number hierarchy on transitive verbs with some combinations of subject and object argument number. Where the object is a human/high animate with PC/PL number, and the subject has SG/DU number, the verb suffix can show agreement with the object in preference to the subject.

[4] Piper describes several processes of deletion or affixation by which the two forms are related.

[5] Piper gives the verb stems *Vgri* and *Vmer* shown in Table 2, where *V* represents an unspecified vowel: the verbs are only recorded with prefixes which result in the phonological deletion of the stem-initial vowel.

Table 2: Verb stems with alternation for NPL vs PL subject

ike	'be(thing).NPL'	*Vgri*	'be(thing).PL'
emri	'sit-down.NPL'	*Vmer*	'sit-down.PL'
akawaret	'climb-onto.NPL'	*etir*	'climb-onto.PL'

tion all come from Group II, and include transitive and intransitive verbs. Table 3 gives examples.

Table 3: Verb stems with alternation for SG/DU vs PC/PL argument

	Intransitive verbs		
ekwey	'stand up.SG/DU.S'	*eko*	'stand up.PC/PL.S'
bakyamu	'go.SG/DU.S'	*bakyaw*	'go.PC/PL.S'
	Transitive verbs		
ep	'carry.SG/DU.O'	*ays*	'carry.PC/PL.O'
diskemer	'chase.SG/DU.O'	*dikes*	'chase.PC/PL.O'
dígwatmu	'pull in from sea. SG/DU.A/O'	*dígwat*	'pull in from sea. PC/PL.A/O'

Piper identifies the number agreement for verb stem alternations of both types (NPL-PL subject, SG/DU-PC/PL object) as being in the main with the subject of intransitive verbs and the object of transitive verbs, in other words showing ergative alignment. However, there are a small number of verb stems that follow different patterns. Piper exemplifies a third type with the verb pair *dígwatmu* 'pull in from sea.SG/DU.A/O' – *dígwat* 'pull in from sea.PC/PL.A/O', shown in Table 3, where the PC/PL verb stem is selected if the either subject or the object has PC/PL number.

A fourth type of distinction is seen with the verb 'to turn over', for which Piper identifies three verb stem variants, conditioned by the number of the subject and/or the object argument: *dipigimer* 'SG/DU.A.turn.over.SG/DU.O', *dipigemeret* 'PC/PL.A.turn.over.SG/DU.O' and *dipiger* 'turn.over.PC/PL.O' For examples see Piper (2013, pp.81-82, exx. 3.10-3.14).

2.1 Anomalies

In the corpus of available texts there are examples where the verb marking does not fit exactly with argument number.[6] In particular, there is a systematic mismatch for some Group II verbs with imperfective aspect marking.

Piper reports that "the present imperfective marker is used to indicate repetition/iterativity" (p.102), adding "forms in the imperfective often involve the paucal/plural form of the verb even when there is no paucal or plural S, A or O". For

[6]This is in addition to the operation of the number hierarchy.

example, for the intransitive verb pair *dígwatmu/dígwat* 'haul in fish', the PC/PL stem must be used in imperfective aspect (1). The SG/DU stem is ungrammatical (2) even if the catch is only one or two fish (Piper, 2013, p.102, ex. 3.79).

(1) *ka* *dígwat-* *li*
1SG.A haul.in.**PC/PL**.fish- PRS.IPFV
'I am hauling in fish.'

(2) **ka* *dígwatmu-* *li*
1SG.A haul.in.**SG/DU**.fish- PRS.IPFV
(Intended) 'I am hauling in a couple of fish.'

Piper also identifies idiosyncratic mismatches between argument number and verb stem. For example, the verb *ker* which generally functions as an auxiliary with the gloss 'do' also has an idiomatic meaning 'fuck'. However, in this case, the PC/PL stem is required even when both arguments to the verb are singular (3).[7]

(3) *ka* *mári* *na-* *ker-* *e*
1SG.A **2SG.O** FUT.1**+1/2.SG.O**- do.**PC/PL.O**- FUT.1
'I will fuck you' (*ibid.* p.193, ex. 5.92)

3 Distinguishing argument number and verbal number

Piper's and Ray's accounts of Meryam Mir contain numerous anomalies and exceptions to argument number agreement. However, assuming the presence of verbal number (Durie, 1986) in the grammar allows the phenomena to be explained more coherently. Durie argues that suppletive verb pairs that are selected by number alternation encode the semantic category of verbal number, rather than syntactic argument number. Grounds for this include: (i) there is a strong tendency for verb pairs to mark "plurality of affect", that is, an absolutive pattern, regardless of a language's syntactic case alignment; (ii) there is no requirement for argument and verbal number marking to coincide, with verbal number agreeing with semantic number where there is a clash; (iii) verbal number can be expressed in languages without argument number marking on nouns or pronouns; and (iv) verbal number can be expressed in contexts where syntactic agreement is not present, such as derivational morphology, or within XCOMP arguments. Durie acknowledges a close relationship between argument and verbal number and concludes, "This provides the potential for concord between verbal number and NP Number, but this concord shows the expected properties of semantic selection rather than agreement." (Durie, 1986, p.365)

As we have seen, Meryam Mir's suppletive verb stem alternations are described by Piper as agreeing in the vast number of cases with either the subject of an intransitive verb or the object of a transitive verb, an absolutive alignment. This

[7] These idiosyncratic patterns may be lexicalised forms, but it is unclear how argument number accounts for the verb stem.

fits with Durie's description of the "plurality of affect" inherent in verbal number. Durie also proposes that verbal number agrees with semantic, not syntactic number, which is arguably the case for examples (1) and (3) above, and that verbal number may mark distinctions not recognised by a language's nouns or pronouns. Meryam Mir in general does not mark argument number on nouns, and there are many examples of mismatch between argument number marked on verb suffixes and marking on the verb stem. The number distinctions marked by verb stem alternation are described by Piper as SG/DU-PC/PL and NPL-PL, neither of which is a distinction marked on nouns or pronouns. Finally, Durie's other condition is that suppletion is seen in contexts where there is no explicit number agreement. The examples of nominalised verbs found in the corpus are derived from verbs that have a single stem form, so it is not possible to test this condition. However, evidence against the other criteria suggests that there is justification for assuming that the suppletive forms mark verbal number.

Accordingly, I conclude that verbal number is a semantic category that can be systematically expressed in Meryam Mir.[8] A summary of the marked number category distinctions and the sites where they are marked is given in Table 4.

Table 4: Available number distinctions and the sites which mark them

				Argument number				Verbal number
Site / Category distinction				Nouns	Pronouns	Verb Prefix/ Infix	Verb Suffix	Verb Root
SG	DU	PC	PL					
No distinction				+	–	–	–	+
SG	NSG			–	+	+	–	–
NPL			PL	–	–	–	–	*
SG/DU		PC/PL		–	–	–	+	*
SG/ PL	DU/PC		SG/ PL	–	–	+	–	–
SG	DU	PC	PL	–	–	–	+	–

* numerous systematic and idiosyncratic anomalies

Arka (2012) assumes that the same feature set underlies argument and verbal number categories in Marori, based on consistency of patterning at the marking

[8]From this point forward, where I am assuming verbal number, I will replace the argument number element of Piper's glosses to reflect this, marking verbal plurality (VPL) or its lack (VSG).

sites. However, Meryam Mir does not show the same consistency of patterning. Verbal number shows only a binary distinction, patterning sometimes with one of two argument number distinctions, either NPL-PL or SG/DU-PC/PL, and sometimes with neither. Furthermore, the NPL-PL distinction is not seen in argument number. Accordingly, I will treat the feature systems for argument number and verbal number separately.

4 Features for argument and verbal number

This discussion of number feature systems follows work by Nordlinger (1997), Dalrymple and Kaplan (2000), Dalrymple et al. (2009), Sadler (2011), Arka (2011, 2012) and others in its treatment of features and agreement, and works within the framework of Lexical Functional Grammar (Bresnan, 2001; Dalrymple, 2001).

In this framework, categories such as case, number, gender, etc are analysed within the f-structure, rather than the c-structure. Category values are determined by unifying the set of features that are contributed from each lexical item in the c-structure that carries a feature specification. Within the c-structure, an individual lexical item can carry the full feature specification for a particular category, be partially-specified or underspecified for a category or feature, or even carry no specification. The final value of a category within an f-structure satisfies the constraints on features and values that are contributed by all lexical items that map to that f-structure. Agreement occurs where the intersection of feature sets carries no conflicting feature values, so the feature set {[+SG][+PL]} is allowed within the grammar. Conversely, if a feature set carries conflicting values, such as {[+SG][−SG]}, it is ungrammatical because of a feature clash.

Under these assumptions, a number feature system for a language has three elements: a set of number categories, each associated with a set of number features and values; feature specifications for each lexical item that is marked for number; and analyses that generate f-structures with the desired meaning for sentences observed in the language, and which can account for ungrammaticality.

Although each language has its own set of available number categories (Corbett, 2000), I assume that the features that generate these categories are drawn from a universal set available to all languages, and that there is a plausible cognitive and semantic motivation for a particular number feature. I also assume that it is preferable for a morphologically-marked binary category distinction to be generated by the alternation of a single feature: if two features are required to vary together, it is unclear why alternations of the single features are not expressed.

4.1 Building on previously proposed feature sets

I will begin by examining proposals on number agreement systems made by Sadler (2011) for Hopi, and Arka (2012) for Marori, assessing the extent to which they can be applied to Meryam Mir, and thus whether a new proposal is necessary.

4.1.1 Sadler's (2011) proposal for Hopi

Sadler looks at Hopi, which has three number categories, SG, DU and PL. Hopi nouns mark all three categories. However, pronouns and verbs mark only SG and PL categories: for pronominal subjects, DU number is expressed by the combination of a PL marked pronoun with a SG marked verb. Sadler accounts for this by proposing a number feature system with two features, [SG] and [PL], which have overlapping entailments, shown in (4)-(5).

(4) $[+\text{SG}] \rightarrow |x| \leq 2$

(5) $[+\text{PL}] \rightarrow |x| \geq 2$

The features combine as shown in Table 5 to generate the three categories.

Table 5: Hopi number features (Sadler, 2011)

	FEATURES	
CATEGORY	[SG]	[PL]
Singular	+	−
Dual	+	+
Plural	−	+

Detailed feature specifications for lexical entries show asymmetries between and within word classes. Nouns are fully specified in all three categories. However, pronouns and verbs are underspecified for one of the two number categories they mark, although for each word class, a different category is underspecified, as shown in Table 6, extracted from Sadler (2011, ex.101, p.412), row labels expanded.

Table 6: Asymmetries in Hopi feature specifications (Sadler, 2011)

		Pronouns	Verbs
Category marking	Singular	$(\uparrow \text{NUM SG}) = +$ $(\uparrow \text{NUM PL}) = -$	$(\uparrow \text{SUBJ NUM SG}) = +$
	Plural	$(\uparrow \text{NUM PL}) = +$	$(\uparrow \text{SUBJ NUM SG}) = -$ $(\uparrow \text{SUBJ NUM PL}) = +$

As a result of this partial underspecification, pronouns marked as "plural" actually indicate number categories DU or PL. Similarly, verbs marked as "singular" actually indicate categories SG or DU.[9] Thus Sadler demonstrates how a sentence consisting only of a pronoun and a verb, neither of which explicitly carry DU marking, can unambiguously represent a DU subject. Examples (6)-(8) are Sadler's examples (p.410, exx 93a.-c.) with amended glosses.

[9] A more precise description of the pronoun and verb category distinctions, which are both given by Sadler as "SG-PL", would be NPL-PL for pronouns and SG-NSG for verbs.

(6) *Pam* *wari*
That.(NUM = SG) run.PFV.(SUBJ NUM = SG ∨ DU)
'S/he ran.'

(7) *Puma* *yúutu*
That.(NUM = DU ∨ PL) run.PFV.(SUBJ NUM = PL)
'They ran.'

(8) *Puma* *wari*
That.(NUM = DU ∨ PL) run.PFV.(SUBJ NUM = SG ∨ DU)
'They (two) ran.'

4.1.2 Arka's (2012) proposal for Marori

Arka's account of Marori, a Papuan language, describes five number categories SG, DU, Limited Plural (LimPL), Generic Plural (GenPL) and Large Plural (LgPL).[10] It has separate morphological marking for argument and verbal number. Arka assumes the same feature set for both argument and verbal number, {[±SG], [±PL], [±AUG]}, where [+AUG] indicates augmentation of the semantic range of a feature. Table 7 shows how the categories are generated from combinations of feature values.

Table 7: Marori number features (Arka, 2012, p.40)

	FEATURES		
CATEGORY	[SG]	[PL]	[AUG]
Singular	+	−	−
Dual	−	−	−
Limited Plural	−	−	+
Generic Plural	−	+	−
Large Plural	−	+	+

From Table 7 the following entailments can be derived.

(9) $[+\text{SG}_{Arka}] \rightarrow |x| = 1$

(10) $[+\text{PL}_{Arka}] \rightarrow |x| \geq 4+$

A comparison with Sadler's features (4, 5), shows that Arka's proposal has a gap in the entailments of the two features, rather than the overlap of entailments which is required to construct the range of Hopi number categories. However, the entailment of $[+\text{SG}_{Arka}]$ is the converse of $[+\text{PL}_{Sadler}]$, which suggests that it may be possible to remove the gap in Arka's entailments by reversing the polarity of the definitions, shown in (11, 12).

[10]The upper boundary of LimPL is not fixed, GenPL entails more than 3 items.

(11) $[+\text{SG}_{Arka}] \rightarrow |x| = 1 \Rightarrow [-\text{SG}_{Arka}] \rightarrow |x| \geq 2 = [+\text{PL}_{Sadler}]$

(12) $[+\text{PL}_{Arka}] \rightarrow |x| \geq 4+ \Rightarrow [-\text{PL}_{Arka}] \rightarrow |x| \leq 3+ \approx [+\text{SG}_{Sadler}]$

If we reverse the polarity of Arka's features such that $[+\text{SG}'_{Arka}]$ = $[-\text{PL}_{Arka}]$ and $[+\text{PL}'_{Arka}]$ = $[\pm\text{SG}_{Arka}]$, we derive the category specifications shown in Table 8.

Table 8: Comparing revised Marori number features with Hopi

	Marori features			Hopi features	
CATEGORY	$[\text{SG}'_{Arka}]$	$[\text{PL}'_{Arka}]$	[AUG]	$[\text{SG}_{Sadler}]$	$[\text{PL}_{Sadler}]$
Singular	+	–	–	+	–
Dual	+	+	–	+	+
Limited Plural	+	+	+	*n/a*	*n/a*
Generic Plural	–	+	–	–	+
Large Plural	–	+	+	*n/a*	*n/a*

There is a discrepancy in the entailments of $[\pm\text{SG}_{Sadler}]$ and $[\pm\text{SG}'_{Arka}]$ (13).

(13) $[+\text{SG}_{Sadler}] \rightarrow |x| \leq 2$; $[+\text{SG}'_{Arka}] \rightarrow |x| \leq 3+$

However, this discrepancy only occurs in Marori for the LimPl category with number values of 3+, which also carry the feature [+AUG]. As Arka defines [+AUG] as augmenting the semantic space of other co-occurring features, this is just what we expect. We can therefore redefine the entailment of the universal feature [±SG] which is sensitive to the feature set F of its category as follows.

(14) $[+\text{SG}] \rightarrow |x| \leq n$; $[+\text{AUG}] \notin F \Rightarrow n = 2$, else *n* set for a given language

With this revised definition, we now have feature systems for Hopi and Marori that are compatible with a universal feature set. [±SG] is defined formally in (14), [±PL], defined formally in (5), and [±AUG] defined informally.

4.1.3 Testing with Meryam Mir

It is now possible to use the feature set elaborated from Sadler's and Arka's proposals to test whether it generates the set of category distinctions observed at different marking sites in Meryam Mir, and also Arka's assumption that the feature sets for argument number and verbal number are identical.

In comparison to Marori, Meryam Mir has only four number categories, and so it is necessary to identify the correspondence between the two. Aggregating

the Marori feature sets is not possible because it would create specifications with inherent feature clash, as shown for one of the possible unifications in (15).[11]

(15) Generic Plural ∪ Large Plural
= {[−SG][+PL][−AUG]} ∪ {[−SG][+PL][+AUG]}
= * {[−SG][+PL][+AUG][−AUG]}

Based on Arka's and Piper's accounts, I assume that the categories SG and DU are identical in Marori and Meryam Mir. This leaves the following three options for mappings between the various Plural categories in Marori and PC/PL in Meryam Mir, from which the feature specifications in Table 9 are derived.

(i) $\text{PC}_{MeryamMir}$ = LimPL_{Marori}
$\text{PL}_{MeryamMir}$ = GenPL_{Marori}

(ii) $\text{PC}_{MeryamMir}$ = LimPL_{Marori}
$\text{PL}_{MeryamMir}$ = LgPL_{Marori}

(iii) $\text{PC}_{MeryamMir}$ = GenPL_{Marori}
$\text{PL}_{MeryamMir}$ = LgPL_{Marori}

Table 9: Possible Meryam Mir number feature specifications mapped from Marori

Category	Feature specifications								
	Mapping (i)			Mapping (ii)			Mapping (iii)		
	[SG]	[PL]	[AUG]	[SG]	[PL]	[AUG]	[SG]	[PL]	[AUG]
Singular	+	−	−	+	−	−	+	−	−
Dual	+	+	−	+	+	−	+	+	−
Paucal	+	+	+	+	+	+	−	+	−
Plural	−	+	−	−	+	+	−	+	+

Looking at the relationship between categories and features for the three mappings, mapping (i) seems intuitively most satisfactory, with Paucal number representing an augmentation of the semantic space of Dual, and Plural being neither singular nor augmented. In mapping (ii), the Plural category is the augmented semantic space of a the bare plural feature, which does not appear. And in mapping (iii), the feature [+PL] without augmentation denotes the Paucal category, which is available only to humans and high animates. This is not impossible, but does raise questions about the universality of the definition for the feature [±PL].

The category distinctions we are seeking to explain were presented in Table 4 above. In Table 10 we see the features whose values are required to alternate, in order to generate each of the observed number category distinctions.

[11] Harbour (2007), in accounting for number patterns in Kiowa, explicitly requires feature clashes, which he describes as 'overspecification', but these cases are then unambiguously marked in the morphosyntax. This does not appear to be the case in Meryam Mir and so I maintain the assumption that feature clashes are ungrammatical.

Table 10: Feature alternations required for argument number distinctions

Category distinction	Mapping (i)	Mapping (ii)	Mapping (iii)
SG–NSG	[PL]	[PL]	[PL]
SG/PL–DU/PC	[SG][PL]	[SG][PL]	[PL][AUG]
SG/DU–PC/PL	[SG][PL][AUG]	[AUG]	[SG]
SG–DU–PC–PL	[SG][PL][AUG]	[SG][PL][AUG]	[SG][PL][AUG]

None of the three mappings is able to generate the binary distinction SG/PL–DU/PC by varying a single feature. Mappings (i) and (ii) denote DU/PC with the feature values [+SG][+PL], whereas for SG/PL, one of those two features has a negative value. Under mapping (iii), DU/PC is observed when the values of [±PL] and [±AUG] are identical, both either positive or negative, and SG/PL is observed when those two features have differing values. Thus none of the three mappings satisfactorily account for the binary distinctions.

Furthermore, under mapping (i), the category distinction SG/DU–PC/PL is derived only by alternating all three number features. It appears therefore that the most plausible mapping of feature specifications, mapping (i), is the most problematic for explaining argument number, and the other two less plausible mappings are also problematic.

For verbal number, the numerous exceptions to argument number agreement suggest a distinct feature set. Even where verbal number marking does align with argument number, the binary NPL-PL distinction described by Piper is only generated by a single feature alternation in mapping (iii), and here it is the feature [±AUG] that varies. As [±AUG] is a secondary feature that extends the semantic range of other features, this is not a satisfactory explanation. Accordingly, I conclude that verbal number alternations are generated by a different feature set.

4.2 Revised proposal: argument number

The three binary distinctions that we are seeking to explain are SG–NSG, SG/DU–PC/PL and SG/PL–DU/PC. The first two of these share the same entailment as the features [±PL] and [±SG] respectively, assuming that the feature [±AUG] is not present and so the entailment of [+SG] is $|x| \leq 2$. However, the distinction SG/PL–DU/PC is not easily delivered. This distinction is therefore a good place to start looking for possible alternative features.

One point that the DU and PC number categories have in common is that, although nouns in Meryam Mir do not routinely mark number, the suffixes *-ba* and *-ey* are optionally available for common and proper nouns. These suffixes indicate membership of a group and trigger DU or PC suffix agreement as appropriate.

Wood (2007) discusses the cognitive process of constructing groups based on

similarity, and how this might apply to the construction of plural actions as well as nominal plurals. A number of features could be proposed that distinguish members of small groups from both individuals and pluralities. One possibility is a feature [±GROUP], which foregrounds the membership of a collection of individuals. Another is its reversed-polarity counterpart [±ATOM], where singular entities and undifferentiated plurals are described as atomic, and groups with a countable number of members are not atomic. This second would be consistent with common nouns not carrying number, but would imply that the marking of nouns with the 'group' and 'dual' suffixes *-ba* and *-ey* is triggered by the absence of atomicity. This would overgenerate in situations where a noun was underspecified for [ATOM], as the absence of [+ATOM] would trigger suffixation. This makes the feature [±GROUP] preferable.

Accordingly, as a first approximation, I will assume that Meryam Mir does not use Arka's proposed feature [AUG], but instead has the feature [±GROUP], which is defined as "a group with a countable number of distinguishable atomic entities". This produces the feature specification for Meryam Mir number categories shown in Table 11, which generates the binary category distinctions by a single feature alternation, as shown in Table 12.

Table 11: Revised proposal of a feature set for Meryam Mir

	Features		
Category	[SG]	[PL]	[GROUP]
Singular	+	–	–
Dual	+	+	+
Paucal	–	+	+
Plural	–	+	–

Table 12: Proposed features underlying argument number category distinctions

Category distinction		Varying feature
(a,b)	SG–NSG	[±PL]
(c)	SG/PL–DU/PC	[±GROUP]
(d)	SG/DU–PC/PL	[±SG]
(e)	SG–DU–PC–PL	[SG][PL][AUG]

4.3 Revised proposal: verbal number

As the motivation for separate verbal number categories arises in part from the lack of complete fit with argument number in Meryam Mir, I will begin this section

by returning to those cases where there is lack of agreement between argument number and the number marked on verb stem, or where there is other, inadequately explained morphology related to verbal number, which are illustrated in examples (16)-(18).

(16) *ka dígwat- li*
1SG.A haul.in.fish.VPL- PRS.IPFV
'I am hauling in fish' [*From* (1), *amended gloss.*]

(17) **ka dígwatmu- li*
1SG.A haul.in.fish.VSG- PRS.IPFV
(Intended) 'I am hauling in a couple of fish' [*From* (2), *amended gloss.*]

(18) *ka mári na- ker- e*
1SG.A 2SG.O FUT.1+1/2.SG.O- do.VPL- FUT.1
'I will fuck you' [*From* (3), *amended gloss.*]

Considering these examples, the following points arise in relation to agreement: there is no feature clash between singular pronouns and either VSG or VPL stems (16)-(18); and the ungrammaticality of (17) suggests that there may be a feature clash between the VSG stem and IPFV aspect.

4.3.1 Boundedness

Cusic (1981) explores the linguistic expression of multiple events and identifies the role of "boundedness" both in grammatical expression (including aspect) and in lexical expression (aktionsart: Agrell, 1908) of event plurality. The relationship between aktionsart, aspect and verbal number is further elaborated by Wood (2007) in developing a typology of pluractionality.

Moens and Steedman (1988) do not specifically refer to aspect, but examine the relationship between lexical category of verbs, tense and aspect. They contrast events with states, and subcategorise events according to their duration (atomic/extended) and according to whether or not the event has a consequence for one of its participants (Table 13). They also show how the nature of an event denoted by a verb may shift between their categories in particular sentential contexts, such as tense/aspect combinations. There are pairwise similarities between Moens and Steedman's event categories and the "lexical aspect" categories proposed by Comrie (1976): point/semelfactive; process/activity; culmination/achievement; and culminated process/accomplishment.

Given the interaction between aspect and verbal number in Meryam Mir, and in the light of Moens and Steedman's, and Wood's analyses of the relationship between aktionsart, aspect and verbal number, I propose to investigate the concept of boundedness in relation to aspect and aktionsart, to see if this can support the definition of a verbal number.

Table 13: Moens & Steedman's subcategorisation of events

	Events		States
	atomic	extended	
+conseq	CULMINATION *recognise, spot, win the race*	CULMINATED PROCESS *build a house eat a sandwich*	*understand love, know resemble*
–conseq	POINT *hiccup tap, wink*	PROCESS *run, swim, walk play the piano*	

We can see the relevance of this to Meryam Mir by considering the analogy between example (16), where VPL appears in the imperfective aspect and the behaviour of the verb 'jump' in English. One jumping event is a point event, whereas the state of 'jumping' assumes iteration and an action in progress. In an appropriate context, a state reading is coerced from a verb that at its base is atomic and –conseq in Moens and Steedman's terms. Moving beyond aspect changes, bringing an adverbial into the sentence context can coerce a culmination ('I jumped out of the window.') and even a culminated process ('I jumped until the floorboards gave way.')

None of these transitions require a change in the verb which is used: 'jump' remains grammatical. However, in Meryam Mir the VSG form of 'jump' is not grammatical with imperfective aspect: the VPL form is required (19).

(19) *ka* *éwpamaret- li*
1SG.S jump.VPL- PRS.IPFV
'I am jumping' (Piper, 2013, p.102, ex.3.78)

There appear to be different dimensions of boundedness. Where the difference between VSG and VPL forms is closely related to argument number, there is boundedness of participants. Telic actions are bounded by their inherent goal, whereas states and atelic actions are unbounded. Atomic events — semelfactives and achievements — are bounded within a very short time, whereas states, activities and accomplishments have a much longer duration. Furthermore it appears that a change in only one of these elements of boundedness is sufficient to require the use of a VPL verb where this is available, even if other elements remain bounded. Thus in (16) and (17), the number of participants — subject and object — remains bounded, but the removal of a time boundary by imperfective aspect triggers the use of the VPL verb stem.

In this model, each verb at its core defines an action or state that has an inherent specification of time-boundedness, telicity, and potentially also of the number

of affected participants (e.g. the difference between 'kill' and 'massacre' in English). Alongside this, a VSG stem can indicate a further level of boundedness of the activity, because of a restricted number of repetitions or participants. Accordingly I propose a feature [BOUNDED] which reflects the presence or absence of this restriction.

When we consider aspect, it is clear that, regardless of aktionsart, imperfective aspect indicates that the action is not yet complete, which suggests that imperfective aspect is inherently [–BOUNDED].

(20) $(\text{ASPECT} = \text{IPFV}) \Rightarrow (\uparrow \text{VNUM BOUNDED}) =_c -$

For verb stem alternations, VSG stems are specified as [+BOUNDED]. This is loosely associated with the argument number feature [+SG], although this association can be overridden.[12] The relationship between the verbal number feature [BOUNDED] and the argument number feature[SG] can be stated as:

(21) $(\uparrow \text{VNUM BOUNDED}) = + \Rightarrow (\uparrow \text{ABS NUM SG}) = +$

VPL stems, however, are unspecified for [±BOUNDED].

One challenge to this association is the differing basic patterns of alignment between absolutive argument number and verbal number: Group I verbs with stem alternations tend to follow the distinction NPL-PL, whereas Group II verbs with stem alternations tend to follow the distinction SG/DU-PC/PL. This requires more investigation. Given that the PC category is only available to humans and high animates, low- and inanimate arguments align with VSG only where they are explicitly SG or DU, across all verbs. For humans and high animates the effective distinction seems to be that in Group I, PC human/high animate arguments align with VSG (and under the current proposal would therefore have the feature [+SG], although this is contradictory to the definition in table 14), whereas in Group II they align with VPL.

5 Feature specifications

If the arguments carry these semantic features, feature specifications can then be proposed for number-marked lexical items within each word class. Table 14 shows the proposed values for all argument-marked items. Each entry in the "Morphosyntactic features" column relates to a group of morphemes that are marked for that feature: morphemes within each group may be further differentiated for tense, aspect, mood etc.

The feature specifications for nouns, pronouns, verb pre-/infixes and Group I verb suffixes are reasonably straightforward, each requiring only one feature to vary. For the SG/PL–DU/PC distinction, agreement with the [GROUP] feature is

[12]The language data are insufficient for a formal representation of the conditions for overriding the association.

Table 14: Fragment of morphological number feature specification

Morphosyntactic features	f-structure features
N-suffix:DU (*-ey*)	$(\uparrow \text{NUM GROUP}) = +$ $(\uparrow \text{NUM SG}) = +$
N-suffix:NSG (*-ba*)	$(\uparrow \text{NUM GROUP}) = +$
PrN:SG	$(\uparrow \text{NUM PL}) = -$
PrN:NSG	$(\uparrow \text{NUM PL}) = +$
V-pre-/infix: SG	$(\uparrow \text{ABS NUM PL}) = -$
V-pre-/infix: NSG	$(\uparrow \text{ABS NUM PL}) = +$
V-pre-/infix: SG/PL	$(\uparrow \text{ABS NUM GROUP}) = -$ $(\uparrow \text{ABS ANIM HUM}) = +$
V-pre-/infix: DU/PC	$(\uparrow \text{ABS NUM GROUP}) = +$ $(\uparrow \text{ABS ANIM HUM}) = +$
V-root: VSG	$(\uparrow \text{VNUM BOUNDED}) = +$
V-suffix(Group I): SG/DU	$(\uparrow \text{SUBJ NUM SG}) = +$
V-suffix(Group I): PC/PL	$(\uparrow \text{SUBJ NUM SG}) = -$
V-suffix(Group II): SG	$(\uparrow \text{SUBJ NUM SG}) = +$ $(\uparrow \text{SUBJ NUM PL}) = -$ $(\uparrow \text{OBJ NUM SG}) =_c +$
V-suffix(Group II): DU	$(\uparrow \text{SUBJ NUM SG}) = +$ $(\uparrow \text{SUBJ NUM PL}) = +$ $(\uparrow \text{OBJ NUM SG}) =_c +$
V-suffix(Group II): PC	$(\uparrow \text{ARG}_x \text{ NUM SG}) = -$ $(\uparrow \text{ARG}_x \text{ NUM GROUP}) = +$ $\left(\begin{array}{l} ((\uparrow \text{ARG}_y \text{ NUM SG}) =_c -) \vee \\ ((\uparrow \text{ARG}_y \text{ NUM GROUP}) =_c -) \end{array} \right)$
V-suffix(Group II): PL	$(\uparrow \text{ARG}_x \text{ NUM SG}) = -$ $(\uparrow \text{ARG}_x \text{ NUM GROUP}) = -$
V-suffix(Group II): IPFV	$(\uparrow \text{VNUM BOUNDED}) = -$

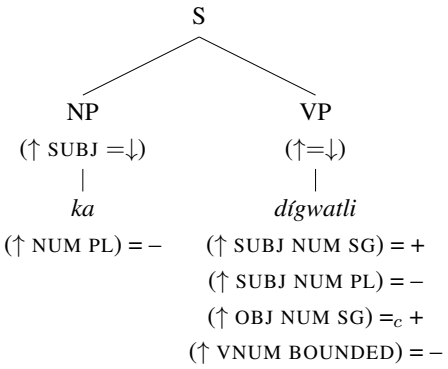
triggered by the presence of a proposed [+HUM] animacy feature, for humans and high animates.

5.1 Feature analysis

The model for verbal number can now be tested by generating c-structures and f-structures for sentences (16)-(18) and seeing whether they correctly predict grammaticality or ungrammaticality. C-structures are set out in examples (22)-(24). For brevity, only those features relating to argument number and verbal number are included in the trees.

Sentence (22) is grammatical as the VPL form of the verb is unspecified for [BOUNDED], and the IPFV suffix contributes the feature [–BOUNDED]. However, there may be a problem arising from the specification of the number suffix. According to the specification given in Table 14, there is a constraint on the object number (↑ OBJ NUM SG) $=_c$ + . Piper does not give examples of the verb with PC/PL object so it is not clear whether the suffix number hierarchy applies. One solution may be that this is an intransitive verb denoting an activity, in which case the object number constraint would not apply.

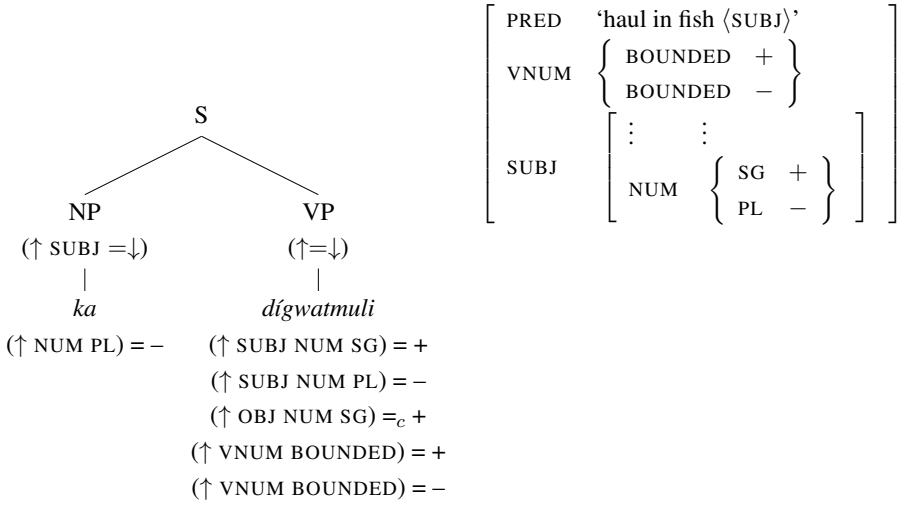
(22) *ka* *dígwat-* *li*
1SG.A haul.in.fish.VPL- PRS.IPFV
'I am hauling in fish'



$$\begin{bmatrix} \text{PRED} & \text{'haul in fish } \langle \text{SUBJ} \rangle \text{'} \\ \text{VNUM} & \left\{ \text{BOUNDED} \quad - \right\} \\ \text{SUBJ} & \begin{bmatrix} \vdots & \vdots \\ \text{NUM} & \left\{ \begin{matrix} \text{SG} & + \\ \text{PL} & - \end{matrix} \right\} \end{bmatrix} \end{bmatrix}$$

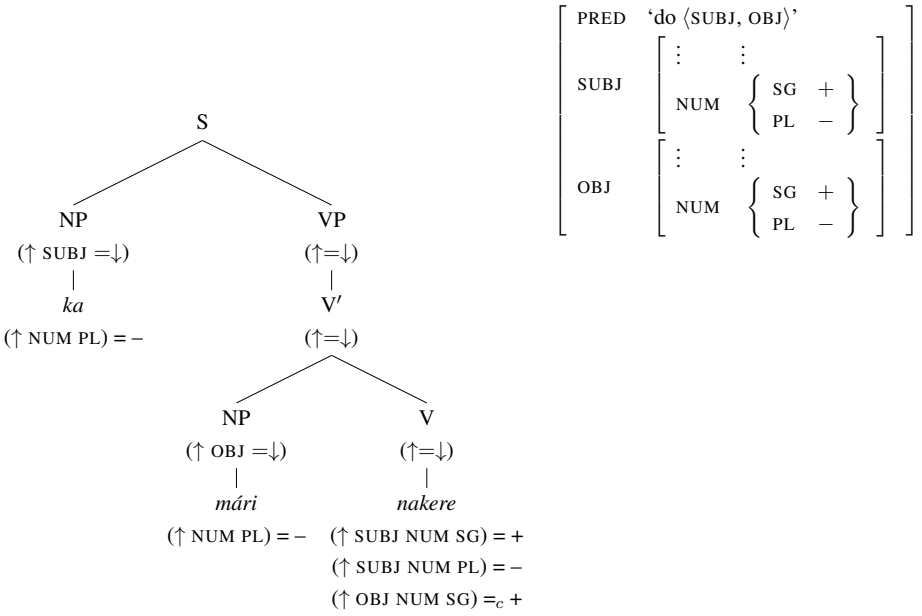
Sentence (23) is ungrammatical as expected, as the inflected verb *dígwatmuli* carries both [+BOUNDED] and [–BOUNDED] features.

(23) **ka* *dígwatmu-* *li*
1SG.A haul.in.fish.VSG- PRS.IPFV
(Intended) 'I am hauling in a couple of fish'



Sentence (24) is grammatical as expected: the VPL form of the verb is unspecified for [BOUNDED] and so there are no feature clashes.

(24) *ka* *mári* *na-* *ker-* *e*
1SG.A 2SG.O FUT.1+1/2.SG.O- do.VPL- FUT.1
'I will fuck you'



6 Conclusions

In conclusion, verb stem alternations in Meryam Mir behave in line with the description of verbal number provided by Durie (1986). However, using verbal number in an account of the language requires distinction of the number features that underlie argument and verbal number marking.

It is possible to generate the observed argument number patterns by adding the feature [±GROUP] to features used in LFG accounts of other languages. However, verbal number patterns require other features. One possibility is a feature [±BOUNDED], which signals a change in the type of event denoted by a verb. However, the size of the currently-available corpus does not allow for full testing. Accounting for other verbal number phenomena not treated in this paper is likely to require further features to be proposed.

References

Agrell, Sigurd. 1908. *Aspektänderung und Aktionsartbildung beim polnischen Zeitworte: Ein Beitrag zum Studium der indogermanischen Präverbia und ihrer Bedeutungsfunktion.*. Lund: H. Ohlsson.

Arka, I Wayan. 2011. Constructive Number Systems in Marori and Beyond. In *Proceedings of the LFG11 Conference*, pages 5–25, Stanford, CA.: CSLI Publications.

Arka, I Wayan. 2012. Verbal Number, Argument Number, and Plural Events in Marori. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of LFG12*, pages 23–43, Stanford, CA: CSLI Publications.

Bresnan, Joan. 2001. *Lexical-functional syntax*, volume 16 of *Blackwell textbooks in linguistics*. Oxford: Blackwell.

Comrie, Bernard. 1976. *Aspect: An introduction to the study of verbal aspect and related problems*, volume 2. Cambridge University Press.

Corbett, Greville G. 2000. *Number*. Cambridge University Press.

Cusic, David Dowell. 1981. *Verbal plurality and aspect*. Ph. D.thesis, Stanford University.

Dalrymple, Mary. 2001. *Lexical Functional Grammar*. San Diego: Academic Press.

Dalrymple, Mary and Kaplan, Ronald M. 2000. Feature indeterminacy and feature resolution. *Language* 76(4), 759–798.

Dalrymple, Mary, King, Tracy Holloway and Sadler, Louisa. 2009. Indeterminacy by underspecification. *Journal of Linguistics* 45(1), 31–68.

Durie, Mark. 1986. The Grammaticization of Number as a Verbal Category. In *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society*, pages 355–368, Berkeley, CA: Berkeley Linguistics Society.

Harbour, Daniel. 2007. *Morphosemantic Number: From Kiowa Noun Classes to UG Number Features*. Dordrecht: Springer.

Moens, Marc and Steedman, Mark. 1988. Temporal ontology and temporal reference. *Computational Linguistics* 14(2), 15–28.

Nordlinger, Rachel. 1997. Morphology building syntax: constructive case in Australian languages. In *Proceedings of the LFG97 Conference, San Diego, California.*, CSLI Publications.

Piper, Nick. 2013. *A sketch grammar of Meryam Mir*. Munich: Lincom Europa.

Ray, Sidney Herbert. 1907. Linguistics. In Alfred Cort Haddon (ed.), *Reports of the Cambridge Anthropological Expedition to Torres Straits*, volume III, Cambridge University Press.

Sadler, Louisa. 2011. Indeterminacy, complex features and underspecification. *Morphology* 21(2), 379–417.

Wood, Esther Jane. 2007. *The semantic typology of pluractionality*. Ph. D.thesis, University of California, Berkeley.

EAST CREE GHOST PARTICIPANTS

Marie-Odile Junker and Ida Toivonen
Carleton University

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King (Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

The Cree verbal morphology includes a relational paradigm, characterized by a morpheme *-w-* on the verb. When verbs occur in this form, the presence in the discourse of an extra third person participant is entailed. We will argue that *-w-* introduces an argument which is realized as a pronominal object in the f-structure. We further argue that what has traditionally been called an obviative morpheme (*-im-*) is in fact an alternative realization of relational morphology.

1 Introduction

The LFG architecture models grammar as distinct, simultaneously present grammatical modules (or *levels* or *structures*).[1] Each of these modules has its own formal and theoretical characteristics. The mapping between the modules is defined with functions and relations (Kaplan and Bresnan, 1982; Dalrymple et al., 1995; Dalrymple, 2001). This *Parallel Projection Architecture* or *Correspondence Architecture* (Kaplan, 1989; Asudeh and Toivonen, 2009) allows for mismatches between different grammatical modules or structures (Bresnan, 2001). Mismatches between phonological phrases and syntactic phrases are common. Also, an auxiliary verb can be a c(onstituent)-structurally independent and complete word while not an f(unctional)-structure "word" (a complete functional structure with a PRED feature and dependents). This paper concerns a particularly interesting apparent mismatch between c-structure and interpretation, so-called *relational constructions* in Cree.

East Cree has a relational form that is marked with a *-w-*. It entails the presence of an extra animate third person, in relation to which the action was performed:[2]

(1) pâhkupayiht-â-**w**-e-u.
dry-AI-REL-TS-3
'She dries it (in relation to him/her/them)'

In (2), the verb is not in the relational form and no extra third person is entailed:

(2) pâhkupayiht-â-u.
dry-AI-3
'She dries it.'

The specific role of the extra "ghost" participant depends on the linguistic or extra-linguistic context. Possible interpretations include a possessive relation or a location

[1]The authors would like audiences at Carleton University and LFG15 for helpful feedback. We especially thank Raj Singh, Miriam Butt, Louise Mycock, Mary Dalrymple, Ron Kaplan, Joan Maling and Ash Asudeh. Our greatest debt is to Junker's East Cree informants. Toivonen wishes to thank the Office of the Dean of FASS at Carleton University for conference travel support. Research for this paper has been partially funded by SSHRC grant #435-2014-1199.

[2]List of abbreviations: REL relational, TS theme sign, II inanimate intransitive, AI animate intransitive, TI transitive inanimate, TA transitive animate, DIR direct, OBV obviative, PROX proximate, DEM demonstrative.

by the third person. Crucially, the participant cannot be mentioned with an overt NP in the immediate clause headed by the relational verb. The referent can, however, be expressed as an overt NP in another clause or as a possessor within an NP.

We propose that the relational morpheme contributes a third person animate object pronoun to the f-structure of the verb. This pronoun is marked in the c-structure with the relational morpheme. No independent object is allowed in the c-structure, neither as a pronoun nor a full NP. This object picks up its referent from the context and it does not have a typical direct object thematic role such as theme or patient.

In addition to the relational *-w-* verbs, there is another class of verbs that is traditionally referred to as obviative verbs, marked by *-im-*. Below, we will follow previous suggestions in the literature and connect the two classes of verbs as different morphological realizations of the relational form.

2 Brief background on East Cree

East Cree is an Algonquian language spoken on the eastern coast of James Bay, in Northern Québec, Canada. It is part of the Cree-Innu dialect continuum described in MacKenzie (1981). There are around 13,000 East Cree speakers in 9 communities. The speakers divide into two major dialects, the Northern and the Southern dialect.

East Cree is a non-configurational (Hale, 1983), polysynthetic (Sapir, 1921), head-marking (Nichols, 1986) language. Cree morphology divides verbs into four distinct classes (Bloomfield, 1946, a.o). The classes are listed in (3) and exemplified in (4):

(3)

INTRANSITIVE VERB WITH INANIMATE SUBJECT	(II)
INTRANSITIVE VERB WITH ANIMATE SUBJECT	(AI)
TRANSITIVE VERB WITH INANIMATE OBJECT	(TI)
TRANSITIVE VERB WITH ANIMATE OBJECT	(TA)

(4)

II	*wâpân*	'it is dawn'	TI	*wâpahtam*	'she sees it'
AI	*wâpû*	'she sees'	TA	*wâpameu*	'she sees him'

Obviation is a central morphosyntactic feature in Algonquian languages (Goddard, 1990): all but one third person participant is marked as a non-topic in a discourse span (Russell, 1991, 1996). The non-topic nouns will receive obviative marking, and only one third person NP can be proximate. Obviation is illustrated by the examples in (5):

(5) a. ni-wâpam-â-u-ch napeu-ch
1-see.TA-TS(DIR.1>3)-3-PL man-PROX.PL
'I see men.'

b. wâpam-e-u-ch napeu-h an-ichii
see.TA-TS(DIR.3>3′)-3-PL man-OBV.PL DEM-PROX.PL
awâsh-ich
child-PROX.PL
'The children see men/a man.'

In (5a), there is only one third person argument, and it is PROXIMATE. Example (5b) contains two third person referents, and only one (the subject) is proximate. The object, *man*, carries obviative marking. Wolfart and Carroll (1981) describe the proximate as "one that is close at hand, center of attention in the discussion, the main point of interest and often the person mentioned first". They describe the obviative as "one that is further away, less important in the conversation, perhaps a figure in the background or one mentioned later".

The gloss DIR in (5) stands for DIRECT marking. In direct marking, mapping onto grammatical functions is determined by the following hierarchy:

(6) SECOND PERSON (2)
FIRST PERSON (1)
THIRD PERSON PROXIMATE (3)
THIRD PERSON OBVIATIVE (3′)

Second person outranks first person, first person outranks third person proximate, and third person proximate outranks third person obviative. In direct marking, the hierarchy in (6) aligns with the hierarchy of grammatical functions. When this mapping is violated, the verb is marked with inverse (INV) morphology.

The verb contains morphemes specifying the arguments of the clause and how they act on each other. These morphemes are called *theme signs* (TS). The TS is found on transitive verbs right after the stem, and it varies according to person combinations and for transitive animates, it indicates direction (direct, inverse). For example, the notation TS(DIR.1>3) in (5a) means a theme sign for 'first person acting on third person'; see Baraby and Junker (2014).

These brief notes obviously do not do justice to the complexities of East Cree morphology and syntax, but we hope that the remarks at least give a sense of what the glosses refer to. For more details, see Junker (2015).

Most of the examples in this paper have been collected by Marie-Odile Junker and colleagues during grammar and verb paradigm documentation workshops held with elders and Cree School Board staff members and during individual sessions with speakers over the last 15 years (Junker and MacKenzie, 2015b,a). Some examples cited here have already been presented in Junker (2003b). Unless otherwise specified, the examples are from the southern dialect of East Cree, but the data have also been confirmed by northern dialect speakers. Relational morphology exists also in other varieties of Cree, and examples from those varieties are marked as such.

3 Relational morphemes in East Cree

3.1 The relational morpheme *-w-*

Cree verbs are sometimes marked with a relational morpheme *-w-*, discussed by Howse (1844); Bloomfield (1928); Ellis (1971); Junker (2003b); Cenerini (2014); and others. Bloomfield (1928) coined the term "relational form". The relational paradigm

always takes direct (not inverse) morphology. The *-w-* morpheme introduces a third person participant into the sentence, and that third person is often a possessor:

(7) a. wâpahtam-**w**-e-u u-mûhkumân-iyû.
see.TI-REL-TS-3 3′-knife-OBV
'S/he sees his/her (someone else's) knife.'

b. ni-wâpahtam-**w**-â-n Mary u-mûhkumân
1-see.TI-REL-TS-1 Mary 3-knife
'I see Mary's knife.'

In (7a), the possessor is not expressed with an overt noun. The referent of the possessor is some third person, but crucially not the same as the subject. In (7b), the possessor is overtly expressed with the NP *Mary*.

The third person participant can only be overtly expressed in the clause if it is embedded as a possessor of an NP (7) or as a dependent of another clause. The third person may not be overtly expressed as a dependent of the verb: it is a "ghost" participant. The referent of the third person is retrieved from the extra-clausal context (unless it is a possessor). The referent is often someone previously mentioned in the surrounding discourse, or someone salient in the context. The specific role of the participant varies greatly, but typical interpretations are as a possessor (as mentioned above),"at his/her place" or "but he/she did not". The (b) examples of (8–9) are relational, and the translations have been suggested by Cree speakers.

(8) a. ni-wâpahte-n mistikw.
1-see.TI-1 wood
'I see a stick.'

b. ni-wâpahtam-**w**-â-n mistiku-yû.
1-see.TI-REL-TS-1 wood-OBV
'I see a stick (but she does not)/(over at her place).'

(9) a. ni-mîyeyihte-n e nipaa-t an awâsh
1-be.happy.TI-1 CONJ eat.AI-3 DEM.PROX.SG child.PROX.SG
'I am happy (that) that child is sleeping.'

b. ni-mîyeyiht-am-**w**-ân e mîchisu-yich ut-awâshim-h
1-like-TI-REL-TS-1 CONJ eat.AI-3′ 3-child-OBV
'I am happy (in relation to her) that her child is eating.'

Below we list an additional example from Innu (Montagnais), where the relational morpheme is *-u-*, and the ghost participant in (10b) and (10c) is Peter.[3]

[3]The examples in (10) are adapted from Drapeau (2014, 286) and Drapeau (2013). The glosses in (10a–b) are ours.

(10) a. nit-ishkuashuku-n assik^{u} ka-tshi usham-an nipi.
1-be.burned.by.something.AI-1 cauldron CONJ-PST boil.TI-1 water
'I was burned by the cauldron after I boiled water.'

b. nit-ishkuashuk(u)-**u**-a-n ut-assik^{u} Pień ka-tshi ushak
1-be.burned.by.sth.AI-REL-TS-1 3-cauldron Peter CONJ-PST boil.TI-1
nipi-ńu.
water-OBV
'I was burned by the (his) cauldron, after Peter boiled water.'

c. [Peter boiled water and when I come to use the same pot]
assiku-ǹu nit-ishkuâshuk(u)-**u**-â-n
pot-OBV 1-be.burned.by.something.AI.REL-TS-1
'I am burned by the pot.'

The examples above show that relational ghost participants are associated with a range of interpretations, and there is no description in the literature of their exact meaning. Scholars describe the contribution of the relational form simply as the addition of a third person:

> "[...] it registers the presence, in the universe of discourse, of additional third person participants." (Junker, 2003b, 319)

> "The relational form is a way to acknowledge the participation, or at least the presence, of an additional third person which benefits or not from the action." (Cenerini, 2014)

> [The relational form expresses] "the involvement of a party other than actor or subject, who is beneficiary or recipient of the action, or is in some way interested in it" (Ellis, 1971)

Despite the fact that the relational form does not refer to a single semantic role, Cree speakers are remarkably good at assigning a specific interpretation to the ghost participant. This can be compared to the ability of speakers of English and many other languages to assign an interpretation to an unexpressed agent in passive sentences.

Let us return to the inability of the third person participant to be expressed with an overt NP in the clause headed by the relational verb:

(11) a. wâpaht-am-**w**-e-u
see-TI-REL-TS-3/3′
'S/he sees it (someone else's knife).'

b. *Mary ni-wâpaht-am-**w**-â-n.
Mary 1-see-TI-REL-TS-1
(intended: 'I see it, in relation to Mary.')

(12) a. ni-nipâ-**w**-â-n
1-sleep.AI-REL-TS-1
'I sleep (at his/her house).'

b. *Mary ni-nipâ-**w**-â-n.
Mary 1-sleep-AI-REL-TS-1
(intended: 'I sleep, in relation to Mary.')

The relational morpheme can in principle refer to a person named Mary, but no overt NP *Mary* can appear in the clause as the relational participant.

3.2 The so-called obviative morpheme *-im-*

The East Cree *-im-* affix has traditionally been called an obviative marker, but it seems to have the same function as relational *-w-*. We argue that *-im-* and *-w-* should be regarded as the same morpheme as far as the syntax and semantics is concerned, even though the two obviously have different phonology, and also different morphology, since they attach to different classes of verbs: while *-w-* goes on intransitive verbs with animate subjects and transitive verbs with inanimate objects, *-im-* goes on transitive verbs with animate objects. The *-im-* morpheme is illustrated in (13)–(14):

(13) a. wâpam-e-u.
see.TA-TS-3
'S/he sees him/her.'

b. wâpam-**im**-e-u
see.TA-REL-TS-3/3′
'She sees him/her (someone else's husband/wife).'

(14) ni-ki paaschisw-**im**-â-yiuh nâ-yiuh wishtaah
1-FUT shoot.TA-REL-TS-3′ DEM-3′ behind.something
'I am going to shoot him, this one (3′) (who is behind this first caribou (3))'

Example (14), adapted from Collette (2014, 158), is from Northern East Cree.

The *-im-* morpheme has been called an obviative morpheme, because it requires third person objects to be obviative. Third person objects can typically be proximate in the appropriate discourse setting, but *-im-* forces the obviative. We propose that the unusual obviative marking in a single animate NP is due to the fact that there is another, more topical, third person participant in the clause, a covert 'ghost' participant.

The *-im-* and *-w-* morphemes share the following traits in common: 1) Both morphemes are used to indicate the presence of an animate third person, a ghost participant. 2) The ghost participant cannot be overtly expressed as a direct dependent of the relational verb. (That is why we call it a ghost participant.) 3) The action denoted by the verb is performed in relation to the ghost participant. 4) Unless the ghost participant is overtly mentioned as a possessor, the referent is picked up from the discourse, or the general context.

Since *-im-* and *-w-* both license the same type of participant, we treat them the same in this paper. The similarity between the two has previously been noted by Junker (2003a,b); Ellis (2004); and Cenerini (2014).

The *-im-* relational morpheme occurs on TA verbs and *-w-* occurs on TI and IA verbs. This leaves (II) verbs (intransitive verbs with inanimate subjects) without a relational form. We have no explanation for this gap, other than it seems consistent with the way animacy is in general respected in Cree: in general, an animate referent will not have a lower grammatical function than an inanimate form. Consider an example like *miywaashuuh* 'they.INAN are beautiful', where the inanimate plural referent is the subject, and the subject is the highest grammatical function. A relational morpheme would add an animate third person participant that would be outranked by the inanimate subject, which is not generally permitted in Cree.

4 Ghost participants: Arguments or adjuncts?

The Cree ghost participants are remarkable in that the verbal morphology forces the presence of a third person in the interpretation of the sentence, but that person may not be expressed as an overt argument of the verb. In typical cases of pronoun incorporation, a full NP and sometimes also a freestanding pronoun can express the same function (in which case the morphologically incorporated pronoun in some cases functions as an agreement marker). Is the ghost participant a true argument of the verb?

4.1 The distinction between arguments and adjuncts

A prototypical argument is a core participant of the verb and syntactically obligatory. The core participant criterion is often considered a semantic requirement, and obligatoriness is instead considered to be a syntactic requirement (see Barbu 2015 for references and discussion). These two criteria do not always line up. For example, the verb *eat* seems to take two core participants, an eater and something eaten. However, *eat* can readily occur without an object in a sentence like *Sue was eating* or *Micky eats early on Fridays*. Also, adverbials are sometimes obligatory, even though they are not core participants of the verb (Goldberg and Ackerman, 2001):

(15) a. These tomatoes grow quickly.
b. ?? These tomatoes grow.

The two basic criteria for argumenthood thus do not always combine to unequivocally determine whether a phrase is an argument or an adjunct, as has previously been noted by Grimshaw (1990); Jackendoff (2001); Koenig et al. (2003); and many others.

A number of syntactic diagnostics for argumenthood have been proposed by, among others, Fillmore (1968); Jackendoff (1977); Pollard and Sag (1987); see Forker (2014) and Needham and Toivonen (2011) for references and discussion. These di-

agnostics neatly distinguish prototypical arguments from prototypical adjuncts.[4] An example argumenthood diagnostic is *the VP anaphora test* (Lakoff and Ross, 1966; Jackendoff, 1977): adjuncts may be added to 'do so' clauses, but arguments may not:

(16) a. Lance broke a chair on purpose and Lambert did so by mistake.
b. *Lance broke a chair and Lambert did so a lamp.

In (16a), *by mistake* is an adjunct and may therefore appear after *did so*. On the other hand, *a lamp* in (16b) is an argument and therefore may not appear after *did so*.

In (16), *by mistake* is a prototypical adjunct, and *a lamp* is a prototypical argument. Not all phrases are, however, so straightforwardly categorized: some classes of phrases seem to fall in between arguments and adjuncts. Instruments, for example, are cross-linguistically difficult to categorize, as they pattern with arguments according to some tests and adjuncts according to others (Donohue and Donohue, 2004; Koenig et al., 2003; Van Valin and LaPolla, 1997). Other types of phrases that seem to fall between arguments and adjuncts include benefactive NPs (e.g., bake *someone* a cake, Toivonen 2013), directional PPs (e.g., run *to the store*, Van Luven 2014), and result XPs (e.g., laugh oneself *silly*, Christie 2013, 2015).

The in-between cases have been treated in various ways in the literature. Kay (2005) calls them *added arguments* and assigns them a specific construction within a Construction Grammar framework. Rákosi (2012) analyzes the in-between cases as *thematic adjuncts*. Asudeh and Giorgolo (2012) argue that different types of dependents differ in how they compose semantically. Manning (2003) argues that the distinction between arguments and adjuncts is simply gradient. Hedberg and DeArmond (2009) distinguishes between primary and secondary arguments.

We make use of Needham and Toivonen's (2011) proposal here. They argue that in-between cases are added or *derived arguments*. These arguments are not part of the verb's basic argument structure; instead, they are licensed by some mechanism of the grammar or lexicon. This mechanism is often but not always tied to derivational morphology. In their LFG account, they specifically propose that derived arguments are arguments licensed by lexical rules in a(rgument)-structure. These "rules" are statements of possible correspondences such as those proposed by Bresnan and Kanerva (1989); Alsina and Mchombo (1989); Bresnan and Moshi (1990); Bresnan and Mchombo (1995). The rules constrain the correspondences between lexical entries in LFG's a-structure, which connects information between the lexicon and f-structure .

4.2 Cree ghost participants as derived arguments

We propose that Cree ghost participants are pronominal arguments, introduced by the relational morphemes *-w-* and *-im-*. This conclusion is not obvious, as these elements display mixed characteristics of arguments and adjuncts, which we will show below.

[4]To our knowledge the so-called argumenthood tests have mostly been applied to English. See, however, Prytz (2012) and Toivonen (2012) for application of the tests to Swedish.

We start our discussion with two adjunct-like characteristics, and then turn to three criteria that indicate that the ghost participants do, in fact, pattern with arguments.

Many regular syntactic argumenthood tests, such as for example the VP anaphora test mentioned above, cannot be applied to the Cree ghost participants, because they are not overtly realized in the clause.[5] There are, however, a few other criteria that apply.

The *core participant* criterion is a basic diagnostic for argumenthood. If an element is semantically required for the event denoted by the verb to take place, it is an argument. In other words, a participant entailed by the verb is an argument. For example, an *eating* event requires an eater and something that is eaten, and a *sending* event requires a sender, something being sent and a recipient. Consider (17):

(17) a. ni-nipâ-n
1-sleep.AI-1
'I sleep.'

b. ni-nipâ**-w-**â-n
1-sleep.AI-REL-TS-1
'I sleep (at his/her house).'

The verb *sleep* entails one core participant, a sleeper. When *sleep* is used with relational morphology, it is understood that an extra third person is present or relevant. However, this extra third person is not part of the core meaning of *sleep*. By the core participants test, then, the ghost participant is an adjunct.

Another test for argumenthood is the *optionality* criterion. According to this criterion, optional phrases are more likely to be adjuncts, and obligatory phrases are arguments. The ghosts are not obligatory; in fact, they are not even optionally present. If these elements are arguments, they are extraordinary arguments indeed.

Prototypical arguments are not optional. However, there are many examples of optional phrases that are nonetheless considered to be arguments, for example, the object of verbs like *wash* and *drink*. When a participant is not expressed (like our ghost participants), the *specific referent* can be used to test its argumenthood status. Forker (2014); Croft (2001); and Matthews (1981) argue that unexpressed canonical arguments are interpreted as referring to some specific person (or people or things): there must be a specific and accessible referent in the discourse context. This is true for the Cree ghost participants. The relational morpheme does not make reference to people in general; it is not interpreted as a generic 'one' or 'they'. By this diagnostic, the ghost arguments thus pattern with arguments, not adjuncts.

East Cree morphology provides two hints about the argumenthood status of the ghost participants. First, the second overt argument in the clause must be obviative:

[5]When they seem to be realized as possessors (e.g., (7b)), the overt possessor is part of the object noun phrase, and the possessor co-refers with the ghost pronominal. We return to this in section 5.

(18) a. ni-wâpam-â-u awâsh.
1-see.TA-TS-3 child
'I see a child.'

b. ni-wâpam-**im**-â-yûh awâsh-a.
1-see.TA-REL-TS-3′ child-OBV
'(John steps back, and when he steps back,) I see a child.'

(19) a. wâpam-e-u u-tem-h.
see.TA-TS-3 3-dog-OBV
'S/he sees his/her (own) dog.'

b. wâpahtam-**im**-e-u u-tem-iyûh.
see.TA-REL-TS-3 3'-dog-OBV
'S/he sees his/her (someone else's) dog.'

In (18a), *awâsh* 'child' is proximate (morphologically unmarked). This is consistent with the requirement of Cree that at most one third person argument is proximate, the others must be obviative. In (15), the subject is a third person proximate and *ute-meh* 'dog' is obviative. The post-verbal NPs in the relational (b) examples both have obviative morphology. This is in principle not problematic: there could be another discourse participant that is more topical or salient and therefore proximate, even if it is not mentioned in the clause. However, what is peculiar is that the postverbal NPs in the (b) examples *must* be obviative. The fact that they cannot be proximate strongly suggests that the ghost participant is an argument of the verb, and it further suggests that this ghost participant is interpreted as being proximate, in the foreground.

Another indication that the ghost participants are arguments of the verb comes from the use of the theme sign. The order of morphemes is given in (20):

(20) PERSON-stem-RELATIONAL-THEME SIGN-PERSON/TENSE/MOOD

AI, TI or TA relational verbs all have transitive animate (TA) theme signs:[6]

(21) a. ni-nipâ-w-**â**-n (AI verb)
1-sleep.AI-REL-TS-1
'I sleep (at his/her house).'

b. ni-wâpahtam-w-**â**-n mistiku-yû. (TI verb)
1-see.TI-REL-TS-1 wood-OBV
'I see a stick (but she does not)/(over at her place).'

c. wâpam-im-**e**-u u-tem-iyûh. (TA verb)
see.TA-REL-TS-3 3'-dog-OBV
'S/he sees his/her (someone else's) dog.'

[6]The *-â-* direct theme sign indicates that a first or a second person acts on a proximate third person. The *-e-* direct theme sign indicates that a proximate third person acts on an obviative third person.

The theme signs included in these verbs are the same that would normally occur only on TA verbs. So, the theme signs behave as if there is an animate object in the clause.

We conclude that the relational participants display mixed argument-adjunct characteristics, but the evidence is consistent with an analysis of the participant as a derived argument.

4.3 Applicatives in East Cree

We argue for an analysis where relational verbs introduce an argument into the clause. This analysis raises the question of how relational verbs differ from applicative verbs. East Cree has a true applicative that is morphologically marked on the verb:

(22) a. chi-kaschisam-**uw**-in.
2-cook.until.tender.TI.2-APPL-(TS)-1/2
'You cook it tender for me.'

b. Mary ni-kaschisam-**uw**-â-u.
Mary 1-cook.until.tender.TI-APPL-TS-3
'I cook it tender for Mary.'

Examples (22a–b) are prototypical applicatives: The verb carries special morphology (*-uw-*), the valency is increased by one, and the new argument is a benefactive.

Relationals are similar to applicatives in that they involve special morphology (*-w-* and *-im-*), and the valency is (arguably) increased by one. However, the relational ghost participants may not be expressed as overt NPs in the clause, and the applicatives differ in this respect. The NP *Mary* in (22b) is an overt realization of the applied argument. Another difference is that the semantic role of the relational participant is not fixed or even restricted. Also, relationals are restricted to third person participants, whereas applicatives are not: the applied object can be of any person and number. For example, in (22a) it is first person singular. Finally, the relational object does not alternate with a corresponding oblique, the way applied objects often do.

We conclude that relationals differ significantly from applicatives. However, they are also similar in that they add an extra participant into the clause, and relationals can therefore be classified as a type of applicative, broadly defined.

5 The syntactic realization of ghost participants

Dahlstrom (1991) presents an LFG analysis of the basic morphosyntax of Plains Cree. We adopt her lexical entries as the basis for our analysis of East Cree relational morphology. The lexical entries and f-structure in (24)–(25) illustrates Dahlstrom's analysis of (23):

(23) ekosi natew awa iskwew ohi kaskite.wastimwa
so fetch.3.OBV this woman this.OBV black.horse.OBV
'So then the woman went and got a black horse.'

(24)

iskwewa, N	(↑ PRED) = ‘woman’ (↑PERS) = 3 (↑GEND) = anim (↑OBV) = −
kaskite.wastimwa, N	(↑ PRED) = ‘black horse’ (↑PERS) = 3 (↑GEND) = anim (↑OBV) = +
awa, D	(↑DEF) = + (↑OBV) = −
ohi, D	(↑DEF) = + (↑OBV) = +
nat-, V_{stem}	(↑ PRED) = ‘fetch ⟨SUBJ OBJ⟩’ (↑OBJ GEND) = anim
-ew, $]_{V\,stem}\_\_]_V$	(↑SUBJ PERS) = 3 (↑SUBJ GEND) = anim (↑SUBJ NUM) = sg (↑SUBJ OBV) = − (↑OBJ PERS) = 3 (↑OBJ GEND) = anim (↑OBJ OBV) = +

(25)

$$\begin{bmatrix} \text{PRED} & \text{‘fetch } \langle \text{SUBJ OBJ} \rangle \text{’} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{‘woman’} \\ \text{NUM} & \text{SG} \\ \text{PERS} & 3 \\ \text{DEF} & + \\ \text{GEND} & \text{anim} \\ \text{OBV} & - \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{‘black horse’} \\ \text{PERS} & 3 \\ \text{DEF} & + \\ \text{GEND} & \text{anim} \\ \text{OBV} & + \end{bmatrix} \end{bmatrix}$$

For discussion of (23)–(25), see Dahlstrom (1991, 122–126).

Recall from section 4 above that the relational morphemes *-w-* and *-im-* introduce an argument. We propose that the relational morphology adds an argument to the argument structure of the verb. Because of the morphological evidence regarding obviative morphology and theme signs (see the discussion in section 4.2), we assume that the argument is an OBJECT. We therefore further propose that the lexical entry of the relational specifies the presence of a third person animate OBJECT (first object) in the f-structure. This object is specified with an obligatory PRED 'pro' feature which prohibits it from co-occurring with an overt object.

Relationals are thus similar to applicatives (section 4.3) and the morphemes are associated with what we will call a relational valency-changing rule in the a-structure. We assume the following rule, which is based on the applicative rule proposed in Bresnan et al. (2016, chapter 14):

(26) Add the relational morpheme *-w-* to IA and TI verbs or add the relational morpheme *-im-* to TA verbs. Add an argument in the second position in the argument structure (i.e. just to the right of the first argument).

Since the argument is added in the second position in the argument structure, all arguments but the first one on the original argument list will be demoted. In the case of TA and TI verbs, the object will be demoted to second object (OBJECT_θ). The rule does not specify the thematic role of the added participant. However, the relational morphemes add the following specifications:

(27)

-im-	(↑ OBJ PRED) = 'pro'		*-w-*	(↑ OBJ PRED) = 'pro'
	(↑ OBJ PERS) = 3			(↑ OBJ PERS) = 3
	(↑ OBJ ANIM) = +			(↑ OBJ ANIM) = +

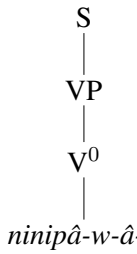
The lexical entries for *-im-* and *-w-* require the object to be third person animate. They also provide a PRED 'pro' feature for these objects.

Lexical entries such as the ones in (27) are familiar from the LFG literature on agreement marking and pro-drop (see Bresnan et al. (2016, chapter 8) for examples, discussion and references). In typical *pro*-drop, the PRED 'pro' feature is optional. When the PRED feature is absent, the morpheme is an agreement marker and the PRED feature of the grammatical function is provided by a syntactically independent NP. When the PRED feature is present, the morpheme is an incorporated pronoun and no independent NP is possible, as that would lead to two PRED features for the same grammatical function (typically a subject or an object). Features can normally unify provided that they are identical. However, each PRED feature takes a unique value and PRED features therefore cannot unify.

The PRED features of the relational morphemes in (27) are obligatory, not optional. The relational objects therefore cannot be overtly expressed as NPs: such NPs would add their own PRED feature and cause a 'PRED clash'. These 'ghosts' can, however, be overtly expressed as dependents of some other head (e.g., an embedded verb or a possessed noun). For other examples of obligatory PRED 'pro' features contributed by the morphology of the head, see Andrews (1990) and Toivonen (2000).

The c-structure and f-structure of *ni-nipâ-w-â-n* (17b) are given in (28):

(28)



$$\begin{bmatrix} \text{PRED} & \text{`sleep } \langle \text{SUBJ OBJ} \rangle \text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`pro'} \\ \text{PERS} & 1 \\ \text{NUM} & \text{SG} \\ \text{ANIMATE} & + \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`pro'} \\ \text{PERS} & 3 \\ \text{ANIMATE} & + \end{bmatrix} \end{bmatrix}$$

Since the lexical entries in (27) specify an animate object, the theme signs that are normally reserved for TA verbs will appear also on AI and TI verbs.[7]

Note that the relational objects are similar to applicative objects cross-linguistically in that applicatives may carry semantic roles that are not prototypical object roles. Prototypical objects are patients or themes (Dowty, 1991), but applied objects can be instruments, for example, as in the Dyirbal example in (29):

(29) a. yabu nguma-nggu balgan yugu-nggu (DYIRBAL)
mother.ABS father.ERG hit stick-INSTR
'Father hit mother with a stick.'

b. yugu nguma-nggu balgalman yabu-gu
stick.ABS father.ERG hit.APPL mother-DATIVE
'Father hit mother with a stick.'

Example (29b) is the applicative version of (29a). The absolutive *yugu* 'stick' in (29b) is an applied object: it is syntactically and morphologically an object. However, its semantic role is still that of an instrument.

Relational objects are even less restricted than objects in applicatives, as the examples above have illustrated. In many cases, the relational morphology merely establishes the existence of an additional third person participant in the context. Consider for example (30) and the corresponding (abbreviated) f-structure in (31):

(30) ni-wâpahtam-w-â-n Mary u-mûhkumân
1-see.TI-REL-TS-1 Mary 3-knife
'I see Mary's knife.'

[7] We could explore an alternative hypothesis where the relational morpheme requires an animate object theme sign, and the theme sign in turn contributes the OBJ agreement features and PRED feature. It seems clear that theme signs can contribute PRED and agreement features: subjects and objects do not need to be overtly expressed, and the information about their characteristics (animacy, etc.) then comes in large part from the theme signs. However, the PRED feature on theme signs must be optional, as theme signs can co-occur with overt NPs.

(31)
$$\begin{bmatrix} \text{PRED} & \text{`see } \langle \text{SUBJ OBJ} \rangle \text{'} \\ \text{SUBJ} & \left[ \text{“I”} \right] \\ \text{OBJ}_i & \begin{bmatrix} \text{PRED} & \text{`pro'} \\ \text{PERS} & 3 \\ \text{ANIMATE} & + \end{bmatrix} \\ \text{OBJ}_\theta & \begin{bmatrix} \text{PRED} & \text{`knife'} \\ \text{POSS}_i & \left[\text{“Mary”} \right] \end{bmatrix} \end{bmatrix}$$

In (30), the relational morpheme introduces a third person into the context. This person is not a possesssor in itself: it takes on the possessor role as a co-referent to the possessor of the knife (in this case *Mary*).

Possessors in relational constructions (30)–(33) raise some interesting questions. In all examples we have with relational morphology and a possessed OBJ_θ, the possessor is co-referenced with the ghost participant. In fact, when a relational sentence includes an object (demoted to OBJ_θ) possessed by a third person, it seems that the possessor is *necessarily* co-indexed by the third person ghost participant, even if another third person referent is available. The pronominal possessor in (32) cannot be interpreted as being coindexed with the subject (contrast with (15) above, which does not include relational morphology).

(32) wâpahtam-**w**-e-u u-mûhkumân-iyû.
see.TI-REL-TS-3 3'-knife-OBV
'S/he sees his/her/their (someone else's) knife.'

(33)
$$\begin{bmatrix} \text{PRED} & \text{`see } \langle \text{SUBJ OBJ} \rangle \text{'} \\ \text{SUBJ}_j & \begin{bmatrix} \text{PRED} & \text{`pro'} \\ \text{PERS} & 3 \\ \text{ANIMATE} & + \\ \text{NUM} & \text{SG} \end{bmatrix} \\ \text{OBJ}_i & \begin{bmatrix} \text{PRED} & \text{`pro'} \\ \text{PERS} & 3 \\ \text{ANIMATE} & + \end{bmatrix} \\ \text{OBJ}_\theta & \begin{bmatrix} \text{PRED} & \text{`knife'} \\ \text{POSS}_i & \begin{bmatrix} \text{PRED} & \text{`pro'} \\ \text{PERS} & 3' \\ \text{ANIM} & + \\ \text{OBV} & + \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

In other words, the relational morphology seems to require disjoint reference between the subject and the possessor. In order to capture this requirement, we could introduce a conditional constraint in the lexical entries for *-w-* and *-im-*: if an OBJ_θ or an ADJUNCT (as in (34) below) has a third person pronominal possessor, that possessor corefers with the relational object. This topic deserves to be investigated further. It is possible that, given the right context, it is possible to interpret (30) as 'I see Mary's knife, in relation to some other third person'. Similarly, in the right circumstances, perhaps (32) can be interpreted as 'She (x) sees some other person's (y's) dog, in relation to yet another person (z).'

(34) nipâ-**w**-e-u u-nipewin-iyi-hch.
sleep.AI-REL-TS-3 3-bed-OBV-LOC
'S/he sleeps in his/her/their (someone else's) bed.'

Another interesting question related to the possessive examples has to do with binding theory. Consider again (30)–(31). The object outranks the possessor in the f-structure; the object "f-commands" the possessor. Depending on which version of binding theory is assumed, this might be problematic according to Principle C, as the possessor R-expression is not free. We set this issue aside for now, but see Bresnan et al. (2016, chapter 10.3) for discussion and references regarding the problematic nature of Principle C.

6 Brief summary

Cree relational morphemes introduce an object pronoun into the f-structure. This pronoun picks up its referent from the linguistic or extra-linguistic context. The only overt realization of the pronoun in the immediate c-structure of the verb is the relational morpheme itself. The reason why the relational object displays mixed argumenthood characteristics is that it is a derived or added argument. It is not a basic argument of the verb, instead it is added by a rule associated with the relational morphemes. This rule is a valency changing rule that is similar to an applicative rule.

References

Alsina, Alex and Mchombo, Sam. 1989. Object asymmetries in the Chicheŵa applicative construction. In Sam Mchombo (ed.), *Theoretical Aspects of Bantu Grammar*, Stanford, CA: CSLI Publications. CSLI Lecture Notes No. 38.

Andrews, Avery. 1990. Unification and morphological blocking. *Natural Language and Linguistic Theory* 8(4), 507–558.

Asudeh, Ash and Giorgolo, Gianluca. 2012. Flexible Composition for Optional and Derived Arguments. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG12 Conference*, pages 64–84, Stanford, CA: CSLI Publications.

Asudeh, Ash and Toivonen, Ida. 2009. Lexical-Functional Grammar. In Bernd Heine and Heiko Narrog (eds.), *The Oxford Handbook of Linguistic Analysis*, pages 425–458, Oxford: Oxford University Press.

Baraby, Anne-Marie and Junker, Marie-Odile. 2014. Flexion. Grammaire innue [Innu Grammar, Aimun-Mashinaikan], http://grammar.innu-aimun.ca/fr/grammaire/verbes/conjugaisons/suffixes/suffixes-thematiques-des-verbes-ta/.

Barbu, Roxana-Maria. 2015. *Verbs and participants: Non-linguists' intuitions*. Masters Thesis, Carleton University.

Bloomfield, Leonard. 1928. The Plains Cree language. In *Proceedings of the International Congress of Americanists*, volume 22, pages 427–431, Rome.

Bloomfield, Leonard. 1946. Algonquian. *Linguistic Structures of Native America* 6, 85–129.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Oxford, UK: Blackwell Publishers.

Bresnan, Joan, Asudeh, Ash, Toivonen, Ida and Wechsler, Stephen. 2016. *Lexical-Functional Syntax, Second edition*. Hoboken, NJ: Wiley-Blackwell.

Bresnan, Joan and Kanerva, Jonni. 1989. Locative inversion in Chichewa: A case study of factorization in grammar. *Linguistic Inquiry* 20(1), 1–50.

Bresnan, Joan and Mchombo, Sam. 1995. The lexical integrity principle: Evidence from Bantu. *Natural Language and Linguistic Theory* 13, 181–265.

Bresnan, Joan and Moshi, Lioba. 1990. Object asymmetries in comparative Bantu syntax. *Linguistic Inquiry* 21, 147–85.

Cenerini, Chantale. 2014. *Relational verbs: Paradigm and practice in a Manitoba dialect of Swampy Cree*. Masters Thesis, University of Regina.

Christie, Elizabeth. 2013. Result XPs and the Argument-Adjunct Distinction. In Butt and King (eds.), *Proceedings of the LFG13 Conference*, pages 212–231, Stanford: CSLI.

Christie, Elizabeth. 2015. *The English Resultative*. Ph. D.thesis, Carleton University.

Collette, Vincent. 2014. *Description de la morphologie grammaticale du cri de l'Est (dialecte du Nord, Whapmagoostui)*. Ph. D.thesis, Université Laval.

Croft, William. 2001. *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.

Dahlstrom, Amy. 1991. *Plains Cree Morphosyntax*. Garland Publications.

Dalrymple, Mary. 2001. *Lexical Functional Grammar. Syntax and Semantics 34*. New York, NY: Academic Press.

Dalrymple, Mary, Kaplan, Ronald M., Maxwell, III, John T. and Zaenen, Annie (eds.). 1995. *Formal Issues in Lexical-Functional Grammar*. Stanford, CA: CSLI Publications.

Donohue, Cathryn and Donohue, Mark. 2004. On the special status of instrumentals. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG04 Conference*, pages 209–225, Stanford, CA: CSLI Publications.

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3), 547–619.

Drapeau, Lynn. 2013. Relational verbs in Montagnais (Innu). In *45th Algonquian Conference*, University of Ottawa.

Drapeau, Lynn. 2014. *Grammaire de la langue innue*. Chicoutimi: Presses de l'université du Québec.

Ellis, C. Douglas. 1971. Cree verb paradigms. *International Journal of American Linguistics* 37(2), 76–95.

Ellis, C. Douglas. 2004. *Spoken Cree ê-ililîmonâniwahk. Level II, West Coast of James Bay*. Edmonton: The University of Alberta Press.

Fillmore, Charles. 1968. The case for case. In Emmon W. Bach and Robert T. Harms (eds.), *Universals in Linguistic Theory*, pages 1–88, New York: Holt, Rinehart and Winston.

Forker, Diana. 2014. A canonical approach to the argument/adjunct distinction. *Linguistic Discovery* 12(2), 27–40.

Goddard, Ives. 1990. Aspects of the topic structure of Fox narratives: Proximate shifts and the use of overt and inflectional NPs. *International Journal of American Linguistics* 56, 307–329.

Goldberg, Adele and Ackerman, Farrell. 2001. The pragmatics of obligatory adjuncts. *Language* 77(4), 798–814.

Grimshaw, Jane. 1990. *Argument Structure*. Cambridge, MA: MIT Press.

Hale, Kenneth. 1983. Warlpiri and the grammar of non-configurational languages. *Natural Language and Linguistic Theory* 1, 5–49.

Hedberg, Nancy and DeArmond, Richard C. 2009. On complements and adjuncts. *Snippets* 19, 11–12.

Howse, James. 1844. *A Grammar of the Cree Language*. London: J.G.F. & J. Rivington.

Jackendoff, Ray. 1977. *X-bar Syntax: A study of phrase structure*. Cambridge, MA: MIT Press.

Jackendoff, Ray. 2001. *Foundations of Language*. Cambridge, MA: MIT Press.

Junker, Marie-Odile. 2003a. East Cree dependent nouns and disjoint reference. *Algonquian and Iroquoian Linguistics* 28(1), 11–13.

Junker, Marie-Odile. 2003b. East Cree relational verbs. *International Journal of American Linguistics* 69(3), 307–29.

Junker, Marie-Odile. 2015. The East Cree language web, http://www.eastcree.org (2000–2015).

Junker, Marie-Odile and MacKenzie, Marguerite. 2015a. The East Cree (Northern Dialect) Verb Conjugation (4th edition), http://verbn.eastcree.org (2010–2015).

Junker, Marie-Odile and MacKenzie, Marguerite. 2015b. The East Cree (Southern Dialect) Verb Conjugation (4th edition), http://verbs.eastcree.org (2011–2015).

Kaplan, Ronald M. 1989. The formal architecture of Lexical-Functional Grammar. In Chu-Ren Huang and Keh-Jiann Chen (eds.), *Proceedings of ROCLING II*, pages 3–18, reprinted in Dalrymple et al. 1995.

Kaplan, Ronald M. and Bresnan, Joan. 1982. Lexical-Functional Grammar: A formal system for grammatical representation. In Joan Bresnan (ed.), *The Mental Representation of Grammatical Relations*, pages 173–281, Cambridge, MA: MIT Press.

Kay, Paul. 2005. Argument-structure constructions and the argument-adjunct distinction. In Hans C. Boas and Mirjam Fried (eds.), *Grammatical Constructions: Back to the Roots*, pages 71–98, Amsterdam/Philadelphia: John Benjamins.

Koenig, Jean-Pierre, Mauner, Gail and Bienvenue, Breton. 2003. Arguments for adjuncts. *Cognition* 89, 67–103.

Lakoff, George and Ross, John R. 1966. Criterion for Verb Phrase Constituency, technical Report NSF-17, Aiken Computation Laboratory, Harvard University.

MacKenzie, Marguerite. 1981. *Towards a Dialectology of Cree-Montagnais-Naskapi*. Ph. D.thesis, University of Toronto.

Manning, Christopher. 2003. Probabilistic syntax. In Rens Bod, Jen Hay and S. Jannedy (eds.), *Probabilistic Linguistics*, pages 289–314, Cambridge, MA: MIT Press.

Matthews, Peter H. 1981. *Syntax*. Cambridge: Cambridge University Press.

Needham, Stephanie and Toivonen, Ida. 2011. Derived arguments. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG11 Conference*, pages 401–421, Stanford, CA: CSLI Publications.

Nichols, Johanna. 1986. Head-marking and dependent-marking grammar. *Language* 62, 56–119.

Pollard, Carl and Sag, Ivan. 1987. *Information-Based Syntax and Semantics*. Stanford: CSLI.

Prytz, Johanna. 2012. In Search of (null) arguments: A review of tests for argumenthood in Swedish. In *N'CLAV conference*, Oslo.

Rákosi, György. 2012. Non-core participant PPs are adjuncts. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG12 Conference*, pages 524–543, Stanford, CA: CSLI Publications.

Russell, Kevin. 1991. Obviation as discourse structure in a Swampy Cree âcimowin. In William Cowan (ed.), *Papers of the 22nd Algonquian Conference*, Ottawa.

Russell, Kevin. 1996. Does obviation mark point of view? In John D. Nichols and Arden C. Ogg (eds.), *Nokotwâsik iskwâhtêm, pâskihtêpayih! Studies in Honour of H.C. Wolfart*, pages 367–382, Winnipeg: Algonquian and Iroquoian Linguistics.

Sapir, Edward. 1921. *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace & World.

Toivonen, Ida. 2000. The morphosyntax of Finnish possessives. *Natural Language and Linguistic Theory* 18(3), 579–609.

Toivonen, Ida. 2012. Between arguments and adjuncts. In *International Conference of Nordic and General Linguistics*, Freiburg.

Toivonen, Ida. 2013. English benefactive NPs. In Butt and King (eds.), *Proceedings of the LFG13 Conference*, pages 503–523, Stanford: CSLI.

Van Luven, Katie. 2014. The argument status of directional PPs, B.A. thesis, Carleton University.

Van Valin, Robert and LaPolla, Randy J. 1997. *Syntax: form, meaning, and function*. Cambridge, UK: Cambridge University Press.

Wolfart, H. Christoph and Carroll, Janet F. 1981. *Meet Cree: A Guide to the Cree Language (second edition)*. Edmonton: The University of Alberta Press.

MODELLING THE SYNTACTIC AMBIGUITY OF THE ACTIVE VS. PASSIVE IMPERSONAL IN LFG

Anna Kibort and Joan Maling
University of Oxford Brandeis University

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King (Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

The *passive* construction, one of the most scrutinised across varying theoretical and typological perspectives, sometimes gives rise to disagreements among linguists about the categorisation of particular cases. Based on data from Irish, Icelandic, Kaqchikel, Polish, and Ukrainian, we argue that so-called 'impersonal passives' are syntactically ambiguous, and can be interpreted in more than one way, as either passives without a subject or as impersonal actives with a null, unspecified, typically human, subject. Transitive impersonals are a key example: even those governing an accusative object may be categorised as either non-promotional passives or impersonal actives. We offer the first LFG analysis of non-promotional passives, and present a way to model the ambiguity between impersonal passives and active impersonals in LFG using Mapping Theory.

1. Introduction

We argue that subjectless impersonal constructions are in principle syntactically ambiguous, and can be analysed as either a non-promotional passive, or an impersonal active with a null, unspecified subject.[1] Several linguists have observed that an intransitive impersonal (i.e. subjectless) construction is inherently ambiguous (Haspelmath 1990: 35; Maling and Sigurjónsdóttir 2002: 126; Blevins 2003: 481). More surprising is the fact that transitive impersonals are syntactically ambiguous in the same way, as shown by the contrasting syntactic behaviour of the accusative-assigning participial *–no/to* construction in two closely-related Slavic languages, Ukrainian and Polish.[2]

(1) a. Polish (Maling and Sigurjónsdóttir 2002, ex. 8b)
*Świątynię zbudowa**no** w 1640 roku.*
church(F).ACC built-**no** in 1640 year
'The church was built in 1640.'

b. Ukrainian (Sobin 1985: 653, ex. 13a)
*Cerkvu bulo zbudova**no** v 1640 roc'i.*
church(F).ACC was built-**no** in 1640 year
'The church was built in 1640.'

Disagreements among linguists about how to analyse such constructions indicate the importance of developing concrete syntactic diagnostics for an active vs. a passive analysis when the direct object shows no signs of promotion to subject, yet there is no

[1] The material in this paper is based in part on work done while the second author was serving as Director of NSF's Linguistics Program. Any opinion, findings and conclusions expressed in this material are those of the authors, and do not necessarily reflect the views of the U.S. National Science Foundation.

[2] The following abbreviations are used in glossed examples: ACC = accusative; CONTEMP = contemporaneous; EXPL = expletive; F = feminine; GEN = genitive; IMPERS = impersonal; INST = instrumental; LOC = locative; M = masculine; MH = masculine human; N = neuter; NMH = non-masculine human; NOM = nominative; PART = participle; PL = plural; PST = past; REFL = reflexive; SG = singular; 1 = first person; 3 = third person.

subject argument expressed on the surface. Cross-linguistically, the syntactic presence of an external argument can be detected in standard ways. For example, a syntactically present subject argument licenses binding of lexical anaphors and control of subject-oriented adjuncts, but blocks an agentive *by*-phrase. Furthermore, unaccusative verbs should be able to occur in the construction, typically with the proviso that the verb selects for a human (internal) argument. A syntactically active impersonal construction with an overt grammatical subject, e.g. French *on* or German *man*, exhibits all of these properties; in contrast, the canonical passive construction lacks all of these properties.[3]

Using these diagnostics, Maling (1993) and Maling and Sigurjónsdóttir (2002: 100–107) contrasted the syntactic properties of the accusative-assigning participial *–no/to* construction in Polish versus Ukrainian. The comparison is summarised in Table 1. This contrast seems puzzling, because in addition to the null subject and non-promoted direct object, both constructions display the same verbal morphology. However, despite their common historical origin, and the shared morphological properties of assigning accusative case and consequent lack of agreement, the Polish and Ukrainian constructions are polar opposites in terms of syntactic behaviour. As Maling and Sigurjónsdóttir document, the Ukrainian *–no/to* construction behaves like a true passive, whereas its Polish counterpart does not (for Polish, see also Kibort 2001; Blevins 2003; Kibort 2004). Note that in addition to the *–no/to* construction, Polish and Ukrainian both have a canonical passive with the expected syntactic properties.

Table 1. Syntactic properties of various constructions in Polish and Ukrainian

syntactic property	Active	Pol/Ukr Passive	Polish *–no/to*	Ukrainian *–no/to*
agentive *by*-phrase	*	ok	*	ok
bound anaphors in object position	ok	*	ok	*
control of subject-oriented adjuncts	ok	*	ok	*
nonagentive ('unaccusative') verbs	ok	*	ok	*

For detailed discussion, see Maling (1993, 2006); Maling and Sigurjónsdóttir (2002: 100–107); Maling and O'Connor (2015), inter alia. The take-home lesson from this comparison is that we cannot tell what the syntactic behaviour of a construction is by looking at superficial morphological properties such as case and agreement. Despite their clearly cognate verbal morphology, Polish and Ukrainian have evidently evolved two syntactically distinct versions of what must have been 'the same' construction at some earlier point. The syntactic properties of the Ukrainian *–no/to* construction show that the

[3] The dichotomy is not always this clear-cut. For example, in German, impersonal passives allow a *by*-phrase, but also reflexives and reciprocals. Both inherent and noninherent reflexive predicates form impersonal passives (see Plank 1993, and especially Schäfer 2012 for discussion); moreover, at least some unaccusative verbs can form impersonal passives (Primus 2011). A Google search turns up examples like *Es wurde auf beiden Seiten gestorben* 'It was died on both sides' containing the unaccusative verb 'to die'. Clearly further investigation of the lexical restrictions is needed. For Icelandic, see Sigurðsson (1989: 322, fn. 48) and Thráinsson (2007: 266ff).

ability to assign accusative case does not necessarily decide between the two possible analyses (contra Haspelmath 1990: 35; Blevins 2003: 481; Danylenko 2006: 262).

The contrasting syntactic behaviour shows definitively that many of the constructions designated as 'non-promotional passives' or 'transitive passives' are actually *impersonal actives*, but also that some are indeed passives according to standard diagnostic syntactic properties. Another case of syntactic change is the so-called *autonomous* construction in Irish, which has been described as a non-promotional passive (Stenson 1989; Noonan 1994). However, McCloskey (2007) argues convincingly that 'the silent subject of an autonomous verb is like an arbitrary subject pronoun, but unlike an implicit agent, in being syntactically active' (p. 828, fn. 3).

2. The Icelandic 'New Impersonal' Construction

A third example of syntactic reanalysis is the New Transitive Impersonal (NTI) construction in Icelandic that has emerged over the past few decades and has occasioned a great deal of disagreement over its categorisation as active or passive. Maling and Sigurjónsdóttir (2002) argue that it is developing into an impersonal active, like the Polish *–no/to* and the Irish autonomous form. However, because the change is ongoing, the evidence is not as categorical as it is for Irish or Polish vs. Ukrainian.

The NTI takes the form in (2); it appears to have a passive participle but differs from the canonical passive in that the verbal object (marked in bold) remains in situ and gets assigned accusative rather than nominative case (if that argument does not bear a lexical case, dative or genitive).

(2) *Loks var fundið* ***stelpuna*** *eftir mikla leit.*
finally was found-N.SG girl.the-F.ACC after great search
'The girl was finally found after a long search.' or
'They finally found the girl after a long search.'

This innovation is a system-internal change that is neither the result of borrowing nor the result of any phonological change or morphological weakening. What exactly is the nature of the change? The analysis of the innovative construction has been the subject of lively debate in recent years; scholars differ in their assessment of whether the NTI is a transitive passive or an active impersonal construction.[4] Everyone agrees that the postverbal NP in the NTI is an object; the disagreement lies in what is assumed to occupy the syntactic subject position. Under one analysis, the NTI is a *non-promotional passive* resembling the Ukrainian participial *–no/to* construction (Eythórsson 2008), and has an empty subject which might be represented as [e]. Under the alternative analysis, the null subject is pro_{arb}, a thematic [+human] subject which can serve as a syntactic binder; i.e. the construction is syntactically *active* like the Polish counterpart (Maling and Sigurjónsdóttir 2002; Maling 2006; Maling and O'Connor 2015).

Icelandic also has a productive impersonal passive of intransitive verbs, which presents an important backdrop to the NTI. The fact that the understood subject of an impersonal passive of an intransitive verb can be interpreted only as a volitional agent (typically human), even if the verb allows inanimate subjects in the active voice, surely

[4] A good survey of the empirical facts and theoretical issues can be found in Thráinsson (2007: 273–283).

supports the plausibility of the pro_{arb} analysis for the New Transitive Impersonal. The subject of the verb *flauta* 'whistle' can be many things, including tea kettles or trains, but the impersonal passive *Það var flautað* 'it$_{EXPL}$ was whistled' can be understood only as describing human whistlers.[5]

The syntactic characteristics of the NTI have been investigated in two nationwide surveys, the first of which was conducted in 1999–2000 and reported in Maling and Sigurjónsdóttir (2002). A questionnaire was distributed to 1,731 tenth graders (age 15–16) in 65 schools throughout Iceland; this number represents 45% of the children born in Iceland in 1984. More than half of the adolescents in most parts of the country (n = 1475) accepted sentences with an accusative definite postverbal object like the one in (2), with a range between 51%–69%, but only 28% of adolescents in Inner Reykjavík (n = 220) accepted these sentences, and very few of the adult controls (n = 200).

A surprising and unexpected result of the survey came from the adult controls. In spite of their disagreements about the syntactic status of the NTI, all scholars of Icelandic considered traditional impersonal passives of intransitive verbs to be true passives. Thus it was a surprise to discover that about half of the adult speakers in the survey accepted two of the diagnostics for *active* constructions – reflexives and subject-oriented adjuncts – in traditional impersonal passives. An example containing a subject-oriented adjunct is shown in (3) (Maling and Sigurjónsdóttir 2002, ex. 37a).

(3) *Það var komið* ***skellihlæjandi*** *í tímann.*
it$_{EXPL}$ was come-N.SG laughing.out.loud into class
'People came into class laughing out loud.'

Maling and Sigurjónsdóttir (2002: 126) pointed out that 'the more subject-oriented participles are accepted, the more simple reflexives are accepted'. For adolescents, the correlation was highly significant ($r = 0.433$, $n = 1693$, $p < 0.001$, 2-tailed); for adults the correlation was also highly significant ($r = 0.532$, $n = 199$, $p < 0.001$, 2-tailed) (Maling and Sigurjónsdóttir 2002: 126, fn. 15, ex. 37a). This correlation supports the suggestion that these speakers have a *syntactically active* representation for the traditional so-called 'impersonal passives'. In contrast, there are other speakers who allow *neither* reflexives *nor* subject-oriented adjuncts; these judgments reflect a *passive* analysis. We take no position on whether the grammar of an individual speaker can have both or only one of the representations. We simply observe that in the aggregate, there is evidence for both grammatical analyses among contemporaneous speakers.

The ongoing syntactic change in Modern Icelandic indicates that native (adult) speakers do not all necessarily come to the same grammatical analysis of every construction; on the contrary, *speakers themselves may come to radically different analyses of the same data*. The readily observable data underdetermines the analysis; it is only by pushing the speaker to judge more complex, or less common (even 'vanishingly rare') sentences that we can see the empirical consequences of choosing one syntactic representation over another. Furthermore, as shown by the independent diachronic developments in Polish, Irish and Icelandic, in which a construction with passive morphology has been reanalysed as a syntactically active construction, the morphosyntactic ambiguity of impersonal constructions can be the locus of syntactic change.

[5] The situation for German and Dutch is more nuanced (see the discussion in Primus 2011). For impersonal passives in Norwegian, see Maling (2006).

3. Kaqchikel

Although we have focused on cases where an apparently passive construction has been reanalysed as an impersonal active, grammatical change can occur in the opposite direction as well (Siewierska 2010, drawing on Broadwell and Duncan 2002). Kaqchikel, an Eastern Mayan language of highland Guatemala, has a variety of passive constructions, including one marked with the suffix *–ki*. The verb in the *ki*-passive shows active morphology, with an active transitive verbal suffix /–Vj/ and the 3rd plural ergative agreement marker *–ki*, as would be appropriate for a transitive verb with an impersonal 'they' subject. Broadwell and Duncan (2002) argue that this verb form has evolved into a construction with the syntactic properties of a passive. It can co-occur with an agentive *by*-phrase, which can be singular or plural, and even 1st or 2nd person. But in contrast with the Ukrainian *–no/to* construction, the *ki*-passive is a promotional passive: it is the patient argument and not the agent which has the grammatical properties of a subject, as shown by syntactic tests including the use of subject-oriented adverbials.

Taken together, our exemplars reveal that every possible association between surface morphology and syntactic behaviour is attested cross-linguistically, as shown in Table 2.

Table 2. Mismatch between morphology and syntax

	Active syntax	Passive syntax
Active morphology	French *on*; German *man*	Kaqchikel *ki*-passive
Passive morphology	Polish *–no/to;* Irish autonomous construction	Ukrainian *–no/to*

In each case that we have discussed above, there are several potential sources of indeterminacy. One is structural: an intransitive impersonal construction is inherently ambiguous (Haspelmath 1990: 35; Maling and Sigurjónsdóttir 2002: 126; Blevins 2003: 481), as are subjectless transitives. As we have seen for Polish, the morphology may be associated historically with a canonical passive even though the syntax indicates an active construction. Because of the inherent morphosyntactic ambiguity, speakers of the same language may construe one of these constructions in different ways, leading to eventual change, as in the Icelandic NTI and in Polish versus Ukrainian.

4. Active impersonal vs. passive impersonal in Mapping Theory (MT)

Mapping Theory is a theory of valency alternations which makes reference to various types of information (see Figure 1). The different levels of representation of a predicate are assumed to be 'linked', with different types of rules mapping one level to another. Figure 1 illustrates a system of mappings between the semantic, lexical, and syntactic representations for a ditransitive predicate such as GIVE.

Fig. 1. Semantic and syntactic valency

referents	<ref$_1$	ref$_2$	ref$_3$>	
instantiated roles[6]	<giver	given	givee>	SEMANTIC/THEMATIC
semantic participants[7]	< *x*	*y*	*b* >	STRUCTURE
	|	|	|	
dependents of the predicate	**<arg$_1$**	**arg$_2$**	**arg$_3$>**	LEXICAL VALENCY
	|	|	|	
grammatical relations	[SUBJ	OBJ	IOBJ]	SYNTACTIC/FUNCTIONAL
syntactic categories	[NP	NP	NP]	SUBCATEGORISATION

The mappings between the semantic structure and lexical valency are referred to as *participant-to-argument mappings*; the mappings between lexical valency and grammatical relations are referred to as *argument-to-function mappings*. Although these terms imply directionality, the mappings result from static constraints and therefore can be understood to apply in both 'directions', i.e. parsing and generation.

4.1. Mapping Theory tools for argument-to-function mappings

A key tool in LFG's Mapping Theory is the decomposition of basic grammatical functions into features. This proposal was originally based on the observation that in argument alternations the arguments which can be identified via their meaning do not have an unrestricted range of options in mapping to grammatical functions. Rather their options are limited to certain grammatical functions from a particular set (see e.g. Bresnan and Kanerva 1989; Bresnan and Zaenen 1990; also Bresnan 2001: 308).

The present version of the Mapping Theory (as developed in particular in Kibort 2007, 2008, 2013, and 2014) uses the same solution, i.e. decomposes basic grammatical functions into features as in the original LFG proposal. However, drawing from another strand of LFG research (see e.g. the overview in Sadler and Spencer 1998), it recognises that there is a substantial difference between meaning-preserving and meaning-altering argument alternations. While the decomposition of argument functions is suitable to model argument structure operations which do not involve a change of meaning, in order to model any meaning-altering argument structure operations the toolbox of the Mapping Theory has to be extended with additional tools. As neither passivisation nor impersonalisation is a meaning-altering operation, this part of the Mapping Theory will not be discussed here in any more detail, but see Kibort (2007 and further work) for a justification of this approach and an application to a wide range of argument alternations.

Thus, basic argument functions[8] are not atomic but arise from particular combinations of more primitive features:

[6] Instantiated roles are specified here only to make the diagram recognisable as an argument structure of the predicate GIVE. The Mapping Theory as proposed here does not need to refer to this level of representation.

[7] See Kibort (2014) for arguments against using generic semantic roles to model participant-to-argument mappings, and extensive references to other works on this subject.

[8] See Kibort (2013) for an overview of the reasons for positing a cross-linguistically motivated grammatical function of the secondary object (OBJ$_\theta$). Beside the four basic functions, many LFG

(4)

	[– r]	[+ r]
[– o]	SUBJ	OBL_θ
[+ o]	OBJ	OBJ_θ

The original LFG interpretation of the features is: [+/–r] thematically (i.e. semantically) restricted; [+/–o] (non)objective. However, in the version of the Mapping Theory assumed here, which preserves a syntactic characterisation of grammatical functions and thereby captures the special status of the secondary object as a 'non-core objective argument', the features are interpreted as follows:

(5) [+o] complements ('internal arguments' of the predicate)
[–o] non-complements (the 'external' argument and oblique arguments)
[–r] core arguments (subject and object only)
[+r] non-core arguments (all arguments except subject and object)

The core of argument structure is a universally available subcategorisation frame which represents the relative syntactic prominence of the arguments of the predicate. This valency template is fixed and the argument positions are characterised by intrinsic features:

(6)

< arg_1	arg_2	arg_3	...	arg_4	...	arg_n >
[–o]/[–r]	[–r]	[+o]		[–o]		[–o]

The ordering of arguments in (6) corresponds to LFG's relational hierarchy of syntactic functions, with adjunct being a non-argument function (Bresnan 2001: 96):

(7) SUBJ > OBJ > OBJ_θ > OBL_θ > COMPL[9] > ADJUNCT

The relational hierarchy is proposed after Keenan and Comrie's (1977) Noun Phrase Accessibility Hierachy, presumed to be universal (at least in nominative-accusative systems):

(8) SUBJ > OBJ > OBJ_θ > OBL > possessor NP > object of comparison

Thus, the ordering of argument positions in (6) also parallels Keenan and Comrie's accessibility hierarchy, however, while LFG's relational hierarchy in (7) is based on final grammatical functions, the ordering in (6) is based on MT's atomic values [+/– r/o].

In the realisation of a particular predicate, the angled brackets contain all and only the selected valency slots for the arguments associated with that predicate, both core and non-core. In other words: predicates do not have to select a contiguous series of arguments. (This can be understood in the sense of the 'derived arguments' of Needham and Toivonen 2011, and is a useful generalisation bearing in mind that the distinction

---

accounts and computational implementations of LFG grammars additionally use COMP and XCOMP for clausal arguments, though some other linguists analyse them as specialised types of the basic grammatical functions (e.g. Zaenen and Engdahl 1994; Alsina, Mohanan and Mohanan 1996; Alsina, Mohanan and Mohanan 2005).

[9] Here, the label COMPL stands for the whole class of various predicate complements (Bresnan 2001: 96).

between arguments and adjuncts is notoriously difficult to justify, see e.g. Przepiórkowski 1999: Ch. 6-10). For example, in *Both parents cooked supper for the children*, the lexical and syntactic valencies of the predicate can be illustrated as follows:[10]

(9) *cook* < arg_1 arg_2 arg_4 >
[–o] [–r] [–o]

The mapping of arguments to grammatical functions follows the Mapping Principle in (10) and the Subject Default in (11):

(10) MAPPING PRINCIPLE
The ordered arguments are mapped to the available functions compatible with their intrinsic marking.

(11) SUBJECT DEFAULT
The first argument compatible with the SUBJ function is mapped to SUBJ.

Note that (10) is a rephrased Mapping Principle. The previous version referred to the 'Markedness hierarchy' of syntactic functions read off the diagram in (4). However, the 'Markedness hierarchy' turns out to be superfluous. Furthermore, since we no longer consider grammatical functions to be 'marked' in the sense originally proposed in Lexical Mapping Theory, the 'Markedness hierarchy' is now also unmotivated. Note also that the Subject Default is not equivalent to the Subject Condition assumed in other variants of LFG's Mapping Theory. Subjectless clauses are permitted (and robustly attested in the world's languages). Only one SUBJ and one OBJ function are permitted by the valency template in (6); however, multiple secondary objects and oblique arguments are possible and distinguished by their subscripts.

Morphosyntactic operations interfere with the default argument-to-function mapping, but do not affect the lexical or semantic levels of representation of the predicate – that is, they are meaning-preserving (see e.g. Sadler and Spencer 1998). Such results are achieved by the mechanism of increasing markedness which preserves monotonicity (Kibort 2007): a morphosyntactic operation can only restrict an argument by adding a 'marked' specification [+r] or [+o] to its syntactic pre-specification. Hence, the available morphosyntactic (i.e. restricting) operations are:

(12) a. adding the [+r] specification to a [–o] argument;
b. adding the [+o] specification to a [–r] argument; and
c. adding the [+r] specification to a [+o] argument.

Each of these operations does not only change the mapping of the grammatical function onto the affected argument, but may also have a knock-on effect on the mapping of grammatical function(s) onto other argument(s).

[10] There is no scope here to discuss the argument/adjunct distinction, but in all examples that follow it is assumed that a non-core semantic participant such as a recipient, instrument, or location, is an argument if it can alternate between an oblique and a core grammatical function.

4.2. The MT model of the morphological impersonal

The morphological impersonal has an argument structure template[11] fully specified for grammatical functions as in the active; (13a) illustrates a transitive active, and (13b) an intransitive active:

(13) a.

	x	*y*	
	|	|	
PREDICATE$_{active}$ ⟨	arg$_1$	arg$_2$	⟩
	[–o]/[–r]	[–r]	
	SUBJ	OBJ	

b.

	x	
	|	
PREDICATE$_{active}$ ⟨	arg$_1$	⟩
	[–o]/[–r]	
	SUBJ	

However, in contrast with the plain active, the morphological impersonal has a 'suppressed' covert SUBJ which can be analysed as a PRO (Kibort 2006, 2008); (14a) illustrates a transitive variant of the Polish *–no/to* impersonal, and (14b) an intransitive variant:

(14) a.

	x	*y*	
	|	|	
PREDICATE$_{\text{-}no/\text{-}to}$ ⟨	arg$_1$	arg$_2$	⟩
	[–o]/[–r]	[–r]	
	SUBJ	OBJ	
	PRO$_{\text{-}no/to}$		

b.

	x	
	|	
PREDICATE$_{\text{-}no/\text{-}to}$ ⟨	arg$_1$	⟩
	[–o]/[–r]	
	SUBJ	
	PRO$_{\text{-}no/to}$	

The introduction of the PRO$_{\text{-}no/to}$ to express the SUBJ of these predicates is realised by the affix *–no/to*. It is this PRO subject which controls subject-oriented adjuncts and binds anaphors and is interpreted as the highest semantic participant of the event. The highest semantic participant has not been downgraded to an oblique, hence it cannot be expressed in a '*by*-phrase'.

4.3. The MT model of the passive

There appears to be a consensus that passivisation is a meaning-preserving argument alternation, i.e. morphosyntactic, as opposed to morphosemantic.

Morphosyntactic operations interfere with the default argument-to-function mapping, but do not affect the lexical or semantic levels of representation of the predicate. Hence, they do not affect the interpretation of the predicate together with its sets of semantic entailments, or the interpretation of the roles of the semantic participants.[12] They affect only the final mapping of grammatical functions to arguments.

[11] An 'argument structure template' is understood here as a generalisation over a set of argument structures of particular predicates. A valency template captures a specific way of mapping from semantic participants to syntactic functions which is the same for all predicates in the set. This concept of valency template can be formalised with the use of LFG's 'templates' (Asudeh, Dalrymple and Toivonen 2008, 2013) and implemented in XLE, as shown by Findlay (2014) and Asudeh, Giorgolo and Toivonen (2014).

[12] As a result, passivisation is considered to preserve the truth conditional meaning component of the active. In contrast, morphosemantic alternations always involve some change of meaning, the

> ‘In general, it is not possible to associate any additional meaning with the passive construction (...), although the overall change in the morphosyntax of the passivized construction usually brings with it at least a shift in the information structure. (...) in [some] cases the passive construction has been co-opted by the grammar to express additional grammatical meanings such as evidentiality (...). However, the typical passive construction is an asemantic rearrangement of subject/object roles.’ (Spencer 2013: 277)

In Mapping Theory, passivisation is an operation which restricts the first argument pre-specified as [–o] (i.e. an unergative argument) by adding to it the [+r] specification (Kibort 2001, 2004). This way, the argument which by default would map to SUBJ is downgraded to the function of an oblique. If there is a second argument which by default would map to OBJ, it has an opportunity to be promoted to SUBJ. The argument structure template in (15a) illustrates the promotional passive of the transitive, and (15b) the (necessarily non-promotional) passive of the intransitive:

(15) a.

	x	*y*	
	|	|	
PREDICATE$_{passive}$ ⟨	arg_1	arg_2	⟩
	[–o]	[–r]	
	[+r]		
	OBL_θ	SUBJ	

b.

	x	
	|	
PREDICATE$_{passive}$ ⟨	arg_1	⟩
	[–o]	
	[+r]	
	OBL_θ	

Note that oblique agents, by virtue of being obliques, are never obligatory. In some languages they may be dispreferred or even unexpressible.

Sentences (16) and (17) exemplify the Polish transitive passive and intransitive passive, respectively:[13]

(16) *Pokój był codziennie sprzątany (przez firmę).*
room(M).NOM was.3SG.M every-day clean.PART.SG.M (by company)
‘The room was cleaned every day (by a company).’

(17) *Było codziennie sprzątane (przez firmę).*
was.3SG.N every-day clean.PART.SG.N (by company)
‘[It] was cleaned every day (by a company). /
There was cleaning every day (by a company).’

Although it is intuitively obvious that (17) is a subjectless passive, there do not seem to be any convenient syntactic tests to prove it. However, this conclusion can be drawn from a particular combination of facts that hold simultaneously of this construction: (i) there is no overt subject, and no positive evidence for a covert syntactic subject; (ii) just like in the transitive personal passive, the agent can be expressed in this construction through an oblique ‘by’-phrase which is not a subject; (iii) just like in the transitive

---

minimal one being some implications which vary between the two alternants (for a compatible view, see Spencer 2013: 294–295 and the reference therein to Pinker 1989).

[13] Out of context, the interpretation of (17) is ambiguous between an intransitive passive (without a subject) and a transitive passive with an omitted *pro*-drop subject, e.g. a dropped 3SG.N noun such as *miejsce* ‘place(N).NOM’ or *pomieszczenie* ‘room(N).NOM’. We are concerned here with the intransitive (subjectless) variant.

personal passive, the agentive oblique – if present – controls adjuncts and anaphors which in the active are controlled by the agentive subject. The pair of sentences in (18) shows that it is the oblique agent in the transitive personal passive, not the subject, which controls the so-called 'subject-oriented' adjuncts; the pair of sentences in (19) shows the same behaviour in the passive of the intransitive:

(18) a. **Pokój był sprzątany oglądając*
room(M).NOM was.3SG.M clean.PART.SG.M watch.$\text{PART}_{\text{CONTEMP}}$
telewizję.
television(F).ACC
'The room was cleaned while watching television.'

cf. b. *Pokój był sprzątany przez nich oglądając*
room(M).NOM was.3SG.M clean.PART.SG.M by them watch.$\text{PART}_{\text{CONTEMP}}$
telewizję.
television(F).ACC
'The room was cleaned by them while watching television.'

(19) a. **Było sprzątane oglądając telewizję.*
was.3SG.N clean.PART.SG.N watch.$\text{PART}_{\text{CONTEMP}}$ television(F).ACC
'There was cleaning while watching television.'

cf. b. *Było sprzątane przez nich oglądając telewizję.*
was.3SG.N clean.PART.SG.N by them watch.$\text{PART}_{\text{CONTEMP}}$ television(F).ACC
television(F).ACC
'There was cleaning by them while watching television.'

The Mapping Theory predicts that the promotional transitive passive, such as the one modelled in (15a), is not the only passive of the transitive that is available. Recall from (12) that all available morphosyntactic (i.e. restricting) operations are:

(20) a. adding the [+r] specification to a [–o] argument;
b. adding the [+o] specification to a [–r] argument; and
c. adding the [+r] specification to a [+o] argument.

Passivisation is an example of an operation which **adds the [+r]** specification to a [–o] argument. In this situation, by default, the second argument pre-specified as [–r] can map to SUBJ. However, another operation may coincide with passivisation: **adding the [+o]** specification to the second, i.e. [–r], argument will force it to be mapped to OBJ and prevent its promotion to SUBJ. Thus, a combination of 'passivisation' and 'object preservation' – simultaneous when viewed from the synchronic perspective – produces a variant of the passive which preserves the structural object:

(21)

	x	*y*	
	|	|	
$\text{PREDICATE}_{\text{pass+obj.pres.}}$ ⟨	arg_1	arg_2	⟩
	[–o]	[–r]	
	[+r]	**[+o]**	
	OBL_θ	OBJ	

The argument structure template in (21) is a model of a non-promotional, impersonal passive of the transitive. This construction allows an oblique agent and does not occur with unaccusatives. As in the 'ordinary' impersonal passive of the intransitive, there is no syntactic subject to bind and control, and the verb has a non-agreeing form.

Sentences (22 a–b) exemplify the Ukrainian *-no/to* construction, which as discussed in section 1, is a non-promotional transitive passive (see Table 1):

(22) Ukrainian (Lavine 2005: 109)

a. *Nemovlja bulo znajdeno u košyku likarjami.*
baby.ACC was found-NO in basket doctors.INST
'A baby was found in a basket by doctors.'

b. *Inozemcja bude posadženo do v'jaznyci hlavoju urjadu.*
foreginer.ACC will-be placed-NO to prison head.INST government
'A foreigner will be put in prison by the head of government.'

Like a (non-promotional) passive of the intransitive, the non-promotional passive of the transitive in Ukrainian does not have a subject that could control subject-oriented adjuncts or bind anaphors.

5. Accounting for ambiguity: syntactic valency frames of predicates

Although the Polish *–no/to* construction is unarguably active and the cognate Ukrainian *–no/to* construction is unarguably passive, the 'new impersonal' construction in Icelandic discussed in section 2 reflects an ongoing change and appears to be in transition from passive to active. The results of the two nationwide surveys, reported in Maling and Sigurjónsdóttir (2002) and Thráinsson et al. (2015), demonstrated clearly that this change was innovated by children. We have argued that what allows this change must be an inherent property of the grammar: non-promoting passives of both intransitives and transitives are inherently ambiguous between a passive and an active impersonal interpretation.

Mapping Theory provides us with argument structure templates which capture the relation between the semantics and the syntax in different constructions. However, if different constructions are not distinguished by their realisation (morphology), surface syntax produces syntactic valency frames which may fit more than one argument structure template, that is, may be a manifestation of more than one distinct construction.

Let us first consider what kinds of syntactic valency frames can be produced by the grammar in a language which has any of the following constructions and their variants: a promotional passive of the transitive (= 'passive'), a non-promotional passive of the transitive (= 'passive + object preservation'), a passive of the intransitive (= 'passive'), a transitive active impersonal (= 'impersonal') and an intransitive active impersonal (= 'impersonal'). The following is a list exemplifying syntactic valency frames of all these kinds of construction. In the interest of clarity, we have limited the examples to those valency frames which contain up to two arguments, and where the arguments express only the two highest semantic participants (furthermore, we have resorted to the traditional semantic role labels in order to facilitate the reading of the examples):

(23) a. transitive

$PREDICATE_{passive}\langle\ SUBJ_{pat/th}\ \ OBL_{ag}\ \rangle$

$PREDICATE_{passive}\langle\ SUBJ_{pat/th}\ \rangle$

. . .

$PREDICATE_{pass+obj.pres.}\langle\ OBJ_{pat/th}\ \ OBL_{ag}\ \rangle$

$\mathbf{PREDICATE_{pass+obj.pres.}\langle\ OBJ_{pat/th}\ \rangle}$

. . .

$\mathbf{PREDICATE_{impers}\langle\ OBJ_{pat/th}\ \rangle}$

. . .

b. intransitive

$PREDICATE_{passive}\ \langle\ OBL_{ag}\ \rangle$

$\mathbf{PREDICATE_{passive}\ \langle\ \ \rangle}$

. . .

$PREDICATE_{passive}\ \langle\ OBL_{ag}\ \rangle$

$\mathbf{PREDICATE_{passive}\ \langle\ \ \rangle}$

. . .

$\mathbf{PREDICATE_{impers}\ \langle\ \ \rangle}$

. . .

It is important to note it is the argument structure templates which enable us to interpret which semantic participant is expressed through which grammatical function for a given predicate.

As should now be clear, in a language which uses the same morphology for any of these constructions (the passive, the passive with a preserved object, and the impersonal), certain syntactic frames with which the predicates in this language are found are necessarily ambiguous. The ambiguous frames are rendered in bold in (23).

One of the ambiguities lies in the syntactic frame of intransitive predicates with no argument present (i.e. the empty frame): this syntactic frame is the result of both the passivisation of an intransitive predicate and the impersonalisation of an intransitive predicate.

The other ambiguity lies in the syntactic frame of transitive predicates with only one argument present, a direct object expressing a patient/theme: this syntactic frame is the result of both the impersonalisation of a transitive predicate and the non-promotional passivisation of a transitive predicate.

If both the active impersonal and the non-promoting passive co-exist in a language and share the same realisation – for example due to changing from one construction to the other – the surface syntax of both unergative transitive and unergative intransitive predicates is identical for the two constructions when the oblique agent is not expressed overtly. Compare:

(24) a. transitive

$PREDICATE_{pass+obj.pres.}\langle\ OBJ_{pat/th}\ \ OBL_{ag}\ \rangle$

$\mathbf{PREDICATE_{pass+obj.pres.}\langle\ OBJ_{pat/th}\ \rangle}$

. . .

$\mathbf{PREDICATE_{impers}\langle\ OBJ_{pat/th}\ \rangle}$

. . .

b. intransitive

$PREDICATE_{passive}\ \langle\ OBL_{ag}\ \rangle$

$\mathbf{PREDICATE_{passive}\ \langle\ \ \rangle}$

. . .

$\mathbf{PREDICATE_{impers}\ \langle\ \ \rangle}$

. . .

Furthermore, even in a language with the familiar (i.e. promoting) passive, when the passive is applied to an unergative intransitive predicate, it results in a surface valency frame identical to the active impersonal of the intransitive when the oblique agent is not expressed. Compare:

(25) a. transitive

$PREDICATE_{passive}\langle\ SUBJ_{pat/th}\ \ OBL_{ag}\ \rangle$

$PREDICATE_{passive}\langle\ SUBJ_{pat/th}\ \rangle$

. . .

$PREDICATE_{impers}\langle\ OBJ_{pat/th}\ \rangle$

. . .

b. intransitive

$PREDICATE_{passive}\ \langle\ OBL_{ag}\ \rangle$

$\mathbf{PREDICATE_{passive}\ \langle\ \ \ \rangle}$

. . .

$\mathbf{PREDICATE_{impers}\ \langle\ \ \ \rangle}$

. . .

As noted earlier, the contrasting syntactic behaviour of the Polish vs. Ukrainian – *no/to* constructions shows that an accusative object is not sufficient to distinguish between the active impersonal and the passive impersonal. The variant of Mapping Theory employed here captures correctly both the different syntactic properties of the constructions in question and the inherent ambiguity of their shared surface syntax.

In the face of syntactic ambiguity it is not surprising that a homophonous non-promoting passive and an active impersonal, as illustrated in (24), may co-exist in a language for a long time. The presence of an oblique agent does distinguish between the active impersonal (which should not allow an oblique agent) and the impersonal passive (which should allow it), but obliques do not have to be expressed and oblique agents are generally not frequent.

6. Modelling grammatical change with Mapping Theory

In section 4 we showed different representations for the impersonal passive and the active impersonal, accounting for their different syntactic behaviour with respect to the subject. In section 5 we showed that the surface syntax of both constructions gives rise to ambiguity which underdetermines the analysis of these constructions. In this section we show how the Mapping Theory can handle the process of grammatical change from the impersonal passive to the active impersonal, and in the opposite direction – with both directions of grammatical change attested, as we will exemplify below.

6.1. Grammatical change from the impersonal *passive* to the *active* impersonal

The oblique agent is a defining property of the passive construction. Therefore, in a language in which there is a way of expressing agents as obliques, the presence of an oblique agent in a syntactic valency frame of a predicate can be taken to indicate that the construction is passive.

However, in the process of grammatical change from the impersonal passive to the active impersonal – as has occurred in the history of Polish, Irish, and is currently occurring in Icelandic – the omission of the oblique agent allows the agent to be re-interpreted as a PRO subject.

In a language with the promoting passive, the locus of the change is the impersonal passive of the intransitive. In principle, predicates which occur in this construction may appear either with an oblique agent or without it. When the oblique agent is not expressed but only implied, predicates in this construction have an empty syntactic valency frame, i.e. no core arguments (SUBJ or OBJ) are ever present with the verbal forms in this construction:

(26) $\text{PREDICATE}_{\text{passive}}$ ⟨ OBL_{ag} ⟩

$\text{PREDICATE}_{\text{passive}}$ ⟨ ⟩

An empty syntactic valency frame is inherently ambiguous between the impersonal passive and the active impersonal interpretation and enables a change of analysis:

(27) **$\text{PREDICATE}_{\text{passive}}$ ⟨ ⟩ ➔ $\text{PREDICATE}_{\text{impers}}$ ⟨ ⟩**

The predicate remains lexically unaltered: it still requires one semantic participant which is interpreted as an agent. In both constructions the agent is unexpressed. The lack of expression of the oblique agent in the passive allows the speaker to formulate a rule in which the implied agent is an unexpressed PRO subject:

(28)

	x			*x*	
	|			|	
$\text{PREDICATE}_{\text{passive}}$⟨	arg_1	⟩	➔	$\text{PREDICATE}_{\text{impers}}$⟨	arg_1 ⟩
	[–o]				[–o]
	[+r]				SUBJ
	(OBL_{ag})				$\text{PRO}_{\text{impers}}$

Grammatical change from the passive to the active may also occur in a language with a non-promoting passive. Furthermore, the non-promoting passive may be an intermediate stage in a language with the promoting passive which innovates the active impersonal from its promoting passive (as has been evident in Icelandic). This stage may facilitate the introduction of the structural object into a construction which is still passive, before the re-interpretation of its agent as a PRO subject switches its analysis to the active.

The following diagram illustrates the change from the promoting to the non-promoting passive of the transitive. In Icelandic, the available evidence shows that the non-promoting passive of the transitive was first innovated in addition to, instead of replacing, the promoting passive of the transitive, for reflexive verbs with the clearly accusative reflexive pronoun *sig* (Árnadóttir et al. 2011), and then for a restricted set of transitive predicates occurring in collocations with particular accusative objects (Sigurðsson 2012). These changes started occurring in parallel with the change of analysis of the impersonal passive of the intransitive. However, as long as the oblique agent phrase was allowed with the accusative objects, we are justified in analysing this construction as a (non-promotional) impersonal passive of the transitive:

(29)

	x	*y*			*x*	*y*	
	|	|			|	|	
$\text{PREDICATE}_{\text{passive}}$ ⟨	arg_1	arg_2 ⟩	➔	$\text{PREDICATE}_{\text{pass+obj.pres.}}$ ⟨	arg_1	arg_2 ⟩	
	[–o]	[–r]			[–o]	[–r]	
	[+r]				[+r]	[+o]	
	(OBL_{ag})	SUBJ			(OBL_{ag})	OBJ	

Again, recall that a syntactic valency frame with an accusative structural object expressing a patient/theme is inherently ambiguous between the non-promotional

impersonal passive interpretation and the active impersonal interpretation, and therefore enables a change of analysis:

(30) **PREDICATE$_{\text{pass+obj.pres.}}$⟨OBJ$_{\text{pat/th}}$⟩ ➔ PREDICATE$_{\text{impers}}$ ⟨OBJ$_{\text{pat/th}}$ ⟩**

The predicate remains lexically unaltered: in addition to a patient/theme participant it still requires a semantic participant which is interpreted as an agent. In both constructions the agent is unexpressed. The lack of expression of the oblique agent in the passive allows the speaker to formulate a rule in which the implied agent is an unexpressed PRO subject:

(31)

	x	*y*			*x*	*y*
PREDICATE$_{\text{pass+obj.pres.}}$ ⟨	arg$_1$	arg$_2$ ⟩	➔	PREDICATE$_{\text{impers}}$ ⟨	arg$_1$	arg$_2$ ⟩
	[–o]	[–r]			[–o]	[–r]
	[+r]	[+o]			SUBJ	OBJ
	(OBL$_{\text{ag}}$)	OBJ			PRO$_{\text{impers}}$	

6.2. Grammatical change from the *active* impersonal to the impersonal *passive*

In the process of grammatical change from the active impersonal to the impersonal passive, an adjunct of cause or reason may be introduced to mean 'because of *x*', initially coindexed with the agentive PRO subject, which means that the cause and the agent have the same referent. The coreferring dependent in this construction may first be expressed by a reflexive. Then the construction may switch to the impersonal passive, with the PRO subject losing its participant status, and the oblique dependent requiring to be expressed through a non-anaphoric element:

(32) **PREDICATE$_{\text{impers}}$ ⟨ ⟩ ➔ PREDICATE$_{\text{passive}}$ ⟨ ⟩**

(33)

	x_i	z_i			*x*	
PREDICATE$_{\text{impers}}$ ⟨	arg$_1$	arg$_4$ ⟩	➔	PREDICATE$_{\text{passive}}$ ⟨	arg$_1$	⟩
	[–o]	[–o]			[–o]	
	SUBJ	(OBL$_{\text{cause}}$)			[+r]	
	PRO$_{\text{impers}}$				(OBL$_{\text{ag}}$)	

The same path is available for the grammatical change from the active to the passive for transitive predicates. One option is for the target construction to be the non-promotional passive of the transitive:

(34) **PREDICATE$_{\text{impers}}$⟨OBJ$_{\text{pat/th}}$⟩ ➔ PREDICATE$_{\text{pass+obj.pres.}}$ ⟨OBJ$_{\text{pat/th}}$ ⟩**

(35)

	x_i	y	z_i			x	y
	|	|	|			|	|
PREDICATE$_{impers}$ ⟨	arg$_1$	arg$_2$	arg$_4$ ⟩	➔	PREDICATE$_{pass+obj.pres.}$ ⟨	arg$_1$	arg$_2$ ⟩
	[–o]	[–r]	[–o]			[–o]	[–r]
	SUBJ	OBJ	(OBL$_{cause}$)			[+r]	[+o]
	PRO$_{impers}$					(OBL$_{ag}$)	OBJ

The other option is for the target construction to be the personal, promotional passive of the transitive, with a patientive subject reanalysed from a topicalised direct object. This path has been argued for Kaqchikel Mayan (Broadwell and Duncan 2002, Broadwell 2006, Siewierska 2010), Kimbundu (Givón 1979), and is common according to Haspelmath (1990). The analysis put forward by these scholars suggests strongly that the non-promotional passive of the transitive, as illustrated in (34) and (35), is a transitional facilitating stage in the grammatical change from the active to the passive in the same way as it is a facilitating stage in the grammatical change in the opposite direction – which was illustrated in (29). To complete the model of the grammatical changes, the following diagram illustrates the change from the non-promoting to the promoting passive of the transitive, i.e. the reverse of (29):

(36)

	x	y				x	y
	|	|				|	|
PREDICATE$_{pass+obj.pres.}$ ⟨	arg$_1$	arg$_2$ ⟩	➔	PREDICATE$_{passive}$ ⟨	arg$_1$	arg$_2$ ⟩
	[–o]	[–r]			[–o]	[–r]
	[+r]	[+o]			[+r]	
	(OBL$_{ag}$)	OBJ			(OBL$_{ag}$)	SUBJ

7. Conclusions

When impersonal passives and active impersonals share the same realisation (morphology), they may be superficially indistinguishable. The impersonal passive of the intransitive may occur without an oblique agent, and if so, its syntactic valency frame is identical to that of the active impersonal of the intransitive. The non-promoting impersonal passive of the transitive may also occur without an oblique agent, and if so, its syntactic valency frame is identical to that of the active impersonal of the transitive. The corollary of this last point is very important: since passives can be non-promotional, an accusative object is not sufficient to distinguish between the passive and the active construction.

In languages in which the two constructions have fully grammaticalised, the passive and the active can be demonstrated to have diametrically different syntactic behaviour. This is due to the fact that the passive does not have a subject, while the active impersonal has a covert PRO subject. Therefore, various language-specific syntactic tests aimed at detecting subjects demonstrate that the two constructions are polar opposites with regard to their subject, even if they have the same or cognate morphology.

In languages in which the constructions are in the process of changing, the tests may not all point to the same conclusion. Since the passive is defined by the alternation in the realisation of the agent argument – as either the subject (in the active) or an omissible

oblique (in the passive) – the presence of an oblique agent in the syntactic valency frame of a predicate does identify that instance of the construction as passive.

In the absence of the context involving subordinate clauses which require control or binding by a subject, the active impersonal and the impersonal passive without an oblique agent remain ambiguous. Syntactic tests regarding control and binding by a subject do usually have the capacity to distinguish between the active impersonal and the passive impersonal in the grammars of particular speakers. However, it is important to remember that both control and binding may be performed by the highest semantic argument – e.g. the agent, or the human (as opposed to non-human) participant – not necessarily by the syntactic subject. This is one of the reasons why the constructions may potentially remain ambiguous even for a single speaker.

The analysis of passives and impersonals presented here is orthogonal to the assumed model of unaccusativity. However, unaccusativity might be the most reliable factor which distinguishes an active impersonal from an impersonal passive without an oblique agent. Our working hypothesis is that only the passive construction should show restricted applicability to a part of the verbal lexicon (even if it correlates with the interpretation of the highest semantic participant as an agent rather than an experiencer), and the active impersonal should be insensitive to unaccusativity (even if it may show a tendency to occur with human agents). This issue remains to be addressed in further work.

The version of the Mapping Theory used in the present work has proven suitable to model the grammatical changes in both directions: from the passive to the active and the opposite. Both changes are enabled by the inherent syntactic ambiguity of the two constructions, and facilitated by small stepwise changes which are predicted by the Mapping Theory.

References

Alsina, Alex, KP Mohanan and Tara Mohanan. 2005. How to get rid of the COMP. In: Butt, Miriam and Tracy Holloway King (eds) *Proceedings of the LFG05 Conference, University of Bergen*. Stanford, CA: CSLI Publications. 21 pp.

Alsina, Alex, Tara Mohanan and KP Mohanan. 1996. Untitled submission to the LFG List. 3 Sept 1996.

Árnadóttir, Hlíf, Thórhallur Eythórsson and Einar Freyr Sigurðsson. 2011. The passive of reflexive verbs in Icelandic. In: Strahan, Tania (ed.) *Nordlyd 37: Relating to Reflexives*, CASTL, Tromsø. 39–97. http://www.ub.uit.no/munin/nordlyd

Asudeh, Ash, Mary Dalrymple and Ida Toivonen. 2008. Constructions with lexical integrity: templates as the lexicon-syntax interface. In: Butt, Miriam and Tracy Holloway King (eds) *Proceedings of the LFG08 Conference, University of Sydney*. Stanford, CA: CSLI Publications. 68–88.

Asudeh, Ash, Mary Dalrymple and Ida Toivonen. 2013. Constructions with lexical integrity. *Journal of Language Modelling* 1(1): 1–54.

Asudeh, Ash, Gianluca Giorgolo and Ida Toivonen. 2014. Meaning and valency. In: Butt, Miriam and Tracy Holloway King (eds) *Proceedings of the LFG2014 Conference, University of Michigan, MI.* Stanford, CA: CSLI Publications. 68–88.

Blevins, James P. 2003. Passives and impersonals. *Journal of Linguistics* 39: 473–520.

Bresnan, Joan. 2001. *Lexical-Functional Syntax.* Oxford: Blackwell.

Bresnan, Joan and Jonni M. Kanerva. 1989. Locative inversion in Chichewa: a case study of factorization in grammar. *Linguistic Inquiry* 20(1): 1–50.

Bresnan, Joan and Annie Zaenen. 1990. Deep unaccusativity in LFG. In: Dziwirek, Katarzyna, Patrick Farrell and Errapel Mejías-Bikandi (eds) *Grammatical Relations. A Cross-Theoretical Perspective.* Stanford, CA: CSLI Publications. 45–57.

Broadwell, G. Aaron. 2006. Passive constructions in Kaqchikel. In: Kulikov, Leonid, Andrej Malchukov and Peter de Swart (eds) *Case, Valency and Transitivity.* Amsterdam: Benjamins. 375–392.

Broadwell, G. Aaron and Lachlan Duncan 2002. A new passive in Kaqchikel. *Linguistic Discovery* 1(2): 1–16.

Danylenko, Andrii. 2006. The 'Greek accusative' vs. the 'new Slavic accusative' in the impersonal environment: an areal or structural discrepancy? In: Danylenko, Andrii *Slavica et Islamica. Ukrainian in Context*. München: Otto Sagner Verlag. 243–265.

Eythórsson, Thórhallur. 2008. The New Passive in Icelandic really is a passive. In: Eythórsson, Thórhallur (ed.) *Grammatical Change and Linguistic Theory: The Rosendal Papers*. Amsterdam: John Benjamins. 173–219.

Findlay, Jamie. 2014. *Mapping Theory without Argument Structure.* Ms. University of Oxford.

Givón, Talmy. 1979. *On Understanding Grammar*. New York: Academic Press.

Haspelmath, Martin. 1990. The grammaticization of passive morphology. *Studies in Language* 14: 25–72.

Keenan, Edward and Bernard Comrie. 1977. Noun phrase accessibility and Universal Grammar. *Linguistic Inquiry* 8(1): 413–461.

Kibort, Anna. 2001. The Polish passive and impersonal in Lexical Mapping Theory. In: Butt, Miriam and Tracy Holloway King (eds) *Proceedings of the LFG01 Conference*. Stanford, CA: CSLI Publications. 163–183.

Kibort, Anna. 2004. *Passive and Passive-like Constructions in English and Polish*. PhD thesis, University of Cambridge.

Kibort, Anna. 2006. On three different types of subjectlessness and how to model them in LFG. In: Butt, Miriam and Tracy Holloway King (eds) *Proceedings of the LFG06 Conference*. Stanford, CA: CSLI Publications. 289–309.

Kibort, Anna. 2007. Extending the applicability of Lexical Mapping Theory. In: Butt, Miriam and Tracy Holloway King (eds) *Proceedings of the LFG07 Conference, Stanford University.* Stanford, CA: CSLI Publications. 250–270.

Kibort, Anna. 2008. Impersonals in Polish: an LFG perspective. In: Siewierska, Anna (ed.) *Impersonal Constructions in Grammatical Theory* [Special Issue of the *Transactions of the Philological Society* 106(2)]. Oxford: Blackwell. 246–289.

Kibort, Anna. 2013. Objects and Lexical Mapping Theory. Abstract. In: Butt, Miriam and Tracy Holloway King (eds) *Proceedings of the LFG2013 Conference, University of Debrecen*. Stanford, CA: CSLI Publications.

Kibort, Anna. 2014. Mapping out a construction inventory with (Lexical) Mapping Theory. In: Butt, Miriam and Tracy Holloway King (eds) *Proceedings of the LFG2014 Conference, University of Michigan, MI.* Stanford, CA: CSLI Publications. 262–282.

Lavine, James E. 2005. The morphosyntax of Polish and Ukrainian *-no/-to. Journal of Slavic Linguistics* 18: 101–130.

Maling, Joan. 1993. Unpassives of unaccusatives. Unpublished handout. Available on ResearchGate.

Maling, Joan, 2006. From passive to active. Syntactic change in progress in Icelandic. In: Lyngfelt, Benjamin and Torgrim Solstad (eds) *Demoting the Agent. Passive, Middle and Other Voice Phenomena*. Amsterdam: Benjamins. 197–223.

Maling, Joan and Sigríður Sigurjónsdóttir. 2002. The 'new impersonal' construction in Icelandic. *Journal of Comparative Germanic Linguistics* 5: 97–142.
Maling, Joan and Catherine O'Connor. 2015. Cognitive illusions: non-promotional passives and unspecified subject constructions. In: van der Zee, Emile et al. (eds) *Structures in the Mind: Essays on Language, Music, and Cognition in Honor of Ray Jackendoff*. Cambridge, MA: MIT Press. 101–118.
McCloskey, James. 2007. The grammar of autonomy in Irish. *Natural Language and Linguistic Theory* 25: 825–857.
Needham, Stephanie and Ida Toivonen. 2011. Derived arguments. In: Butt, Miriam and Tracy Holloway King (eds) *Proceedings of the LFG11 Conference, University of Hong Kong*. Stanford, CA: CSLI Publications. 401–421.
Noonan, Michael. 1994. A tale of two passives in Irish. In: Fox, Barbara A. and Paul J. Hopper (eds) *Voice. Form and Function*. Amsterdam: Benjamins. 279–311.
Pinker, Steven. 1989. *Learnability and Cognition. The Acquisition of Argument Structure.* Cambridge, MA: The MIT Press.
Plank, Frans. 1993. Peculiarities of passives of reflexives in German. *Studies in Language* 17: 135–67.
Primus, Beatrice. 2011. Animacy and telicity: semantic constraints on impersonal passives. In: von Heusinger, Klaus and Helen de Hoop (eds) *Case Variation.* Special Issue of *Lingua* 121: 80–99.
Przepiórkowski, Adam. 1999. *Case Assignment and the Complement/Adjunct Dichotomy. A Non-configurational Constraint-based Approach*. PhD thesis, University of Tübingen.
Sadler, Louisa and Andrew Spencer. 1998. Morphology and argument structure. In: Spencer, Andrew and Arnold M. Zwicky (eds) *The Handbook of Morphology*. Oxford: Blackwell. 206–236.
Schäfer, Florian. 2012. The passive of reflexive verbs and its implications for theories of binding and case. *Journal of Comparative Germanic Linguistics* 15: 213–268.
Siewierska, Anna. 2010. From third plural to passive: incipient, emergent and established passives. *Diachronica* 27: 73–109.
Sigurðsson, Einar Freyr. 2012. *Germyndarleg þolmynd eða þolmyndarleg germynd?* [An active passive or a passive active?] MA thesis, University of Iceland, Reykjavík.
Sigurðsson, Halldór Ármann. 1989. *Verbal Syntax and Case in Icelandic*. PhD thesis, Lund University. Published by the Linguistic Institute, University of Iceland.
Sobin, Nicholas J. 1985. Case assignment in Ukrainian morphological passive constructions. *Linguistic Inquiry* 16(4): 649–662.
Spencer, Andrew. 2013. *Lexical Relatedness. A Paradigm-based Model*. Oxford: Oxford University Press.
Stenson, Nancy. 1989. Irish autonomous impersonals. *Natural Language and Linguistic Theory* 7(3): 379–406.
Thráinsson, Höskuldur. 2007. *The Syntax of Icelandic*. Cambridge: Cambridge University Press.
Thráinsson, Höskuldur, Sigríður Sigurjónsdóttir, Hlíf Árnadóttir and Thórhallur Eythórsson. 2015. Um þolmynd, germynd og *það*. [On Passive, Active, and *there*.] In: Thráinsson, Höskuldur, Ásgrímur Angantýsson & Einar Freyr Sigurðsson (eds) *Tilbrigði í íslenskri setningagerð* II, [*Variation in Icelandic Syntax* II], Chapter 8. Reykjavík: Háskólaútgáfan.
Zaenen, Annie and Elisabet Engdahl. 1994. Descriptive and theoretical syntax in the lexicon. In: Atkins, B. T. Sue and Antonio Zampolli (eds) *Computational Approaches to the Lexicon: Automating the Lexicon II.* Oxford: Oxford University Press. 181–212.

ON NEGATIVE PARTICLES AND NEGATIVE POLARITY IN HUNGARIAN

Tibor Laczkó
University of Debrecen

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King
(Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

In this paper, I modify and augment my LFG-XLE analysis of negation in Hungarian proposed in Laczkó (2014b) by (i) developing an account of the special uses of negative particles (ii) capturing their interaction with negative polarity items (iii) presenting a formal treatment of suppletive negative variants of the copula. In addition, I argue for a particular distribution of labour in my approach for the three standard XLE devices for handling negation phenomena across languages.

1 Introduction

In Laczkó (2014a) I present the basic ingredients of a comprehensive LFG analysis of the preverbal portion of Hungarian finite clauses (designed to be XLE-implementable). I propose a general formal apparatus for handling constituents in the topic and the quantifier fields and in the specifier position of the VP. In Laczkó (2014b) I outline an LFG analysis of constituent and predicate negation in this model. I focus on c-structural, functional and lexical representational issues and leave semantic issues to future research. In this paper I set out to explore these latter issues and present an LFG-XLE treatment by (i) developing an account of the special uses of negative particles, (ii) capturing their interaction with negative polarity items (n-words), (iii) presenting a formal treatment of the two forms of the two suppletive negative variants of the copula. In addition, I argue for a particular distribution of labour in my approach for the three standard XLE devices for handling negation phenomena across languages.

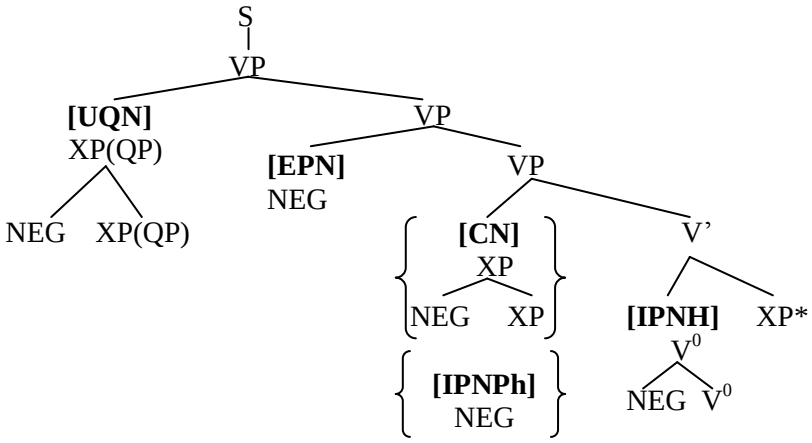
The structure of the paper is as follows. In section 2, I summarize the relevant details of my analysis in Laczkó (2014b). In section 3, I present my generalizations about the behaviour of various types of negative particles and negative polarity items. In section 4, I develop my LFG-XLE analysis of these phenomena. In section 5, I make some concluding remarks.

2 On Laczkó (2014b)

In my analysis, I capitalize on É. Kiss' (1994) structural approach to negation (in her GB framework). Consider the schematic representation of the five major types of negation in (1).[1]

[1] The abbreviations in square brackets indicate the following types of negation: [UQN] = universal quantifier negation, [EPN] = (VP)external predicate negation, [CN] = constituent negation, [IPNPh] = (VP)internal predicate negation, in a phrasal position, [IPNH] = (VP)internal predicate negation, head-adjunction. The curly brackets signal the complementarity of [CN] and [IPNPh].

(1)



A) In my implemented rules, I use the NEG category label (as opposed to Laczkó & Rákosi's (2008-2014) ADV, for instance), which contributes to parsing parsimony to a great extent.

B) I assume NEG to be a uniformly non-projecting word (capable of occurring in both X^0 and XP positions), cf. the treatment of particles in particle-verb-constructions in English, German and Hungarian in Forst et al. (2010), using the category label PRT. My motivation for this is the fact that this Hungarian negative marker does not exhibit any phrasal behaviour in its own right. Most importantly: it can never be modified; thus, there is no empirical evidence for its phrasal projection.[2]

C) As (1) shows, in my analysis NEG can occupy three major types of syntactic positions: it can be in [Spec,VP]: IPNPh, and it can also be either head-adjoined: IPNH or phrase-adjoined: UQN, EPH, CN.

D) In all its uses, it has the ADJUNCT functional annotation in c-structure.

E) I assume the following lexical form for the negative marker.

(2) *nem* NEG (↑ PRED) = 'nem'
(↑ ADJUNCT-TYPE) = neg.

3 Negative particles and negative polarity items

Let me start with an overview of Hungarian pronouns with two sets of examples in Table 1 (next page).

- The first part of the compounds in the first two columns encodes the universal or existential aspect and the second carries the specific pronominal content: 'person, thing, place, etc.'. This second member is typically the corresponding interrogative pronoun in present day Hungarian.

[2] For further discussion, see sections 4 and 5.

- Negative polarity pronouns consist of an allomorph of the *se(m)* negative particle and the usual interrogative pronominal second member.
- They can never occur in a positive clausal environment (as opposed to English negative pronouns): they must always be licensed by a negative particle.
- Negative polarity items are also often called n-words; below I will use this term, and I will also use its acronym: NW.

universal	**existential**	**negative polarity**
MINDEN-	VALA-	SE-
minden-ki every-who 'everybody'	*vala-ki* some-who 'somebody'	*sen-ki* no-who 'nobody'
minden-hol every-where 'everywhere'	*vala-hol* some-where 'somewhere'	*se-hol* no-where 'nowhere'

Table 1. The system of Hungarian pronouns

Consider the examples in (3) and (4), illustrating the basic Hungarian facts.

(3) *János* **(nem)* *látott* *senki-t.*
John.NOM not saw #nobody-ACC
'John didn't see anybody.' or 'John saw nobody.'

(4) *Senki* **(nem)* *látott* *senki-t.*
#nobody.NOM not saw #nobody-ACC
'Nobody saw anybody.'

Given that Hungarian n-words have negative morphological forms, I gloss them with the combination of the hash mark and the corresponding English negative pronouns, e.g. *senki-t* #nobody-ACC, as in (3) and (4). As the customary **(nem)* representation indicates, (3) and (4) are ungrammatical if the negative particle is missing from these sentences. Notice that the negative particle licenses both the n-word preceding it, which is a special case,[3] and the n-word following it, which is the regular situation.

Let me now turn to the types and distribution of negative particles. In addition to the ordinary negative particle *nem* 'not', which we have been dealing with so far, there is another, special particle meaning 'also_not', which has two forms: *sem*[4] and *se*.[5] The two forms have exactly the same

[3] In section 4 I will point out the challenge this poses for a formal analysis, and then I will present my solution.

[4] This variant transparently reflects the relationship between the meaning and the etymology of this particle: ***iS*** 'also' + ***nEM*** 'not' → ***SEM***. É. Kiss (2011) points out that the original forms of the two elements were *es* and *nem*, and they got merged. The former later developed into *is* 'also', an additive particle, and into *és* 'and', a

meaning and distribution, and the only difference between them is that the latter is more informal, and typically it occurs in casual speech. For this reason, I will discuss and represent them jointly by using the *sem* form.

The GB/MP literature on negation in Hungarian in general and on the treatment of negative polarity items in particular is enormous; for a variety of analyses, see É. Kiss (1992, 1994, 2008, 2011, 2015), and the references therein. Therefore, I defer the detailed and systematic comparison of my LFG analysis with several salient GB/MP accounts to another forum.

My basic generalizations about the distribution of the negative particles and negative polarity items are as follows.[6]

- The ordinary negative particle *nem* precedes the constituent that it combines with (by being left-adjoined to it): **nem^XP**),[7] see (1).
- *Sem* is right-adjoined to its respective constituent.
- In addition to its combinability with intrinsic n-words: **NW^sem**, *sem* turns ordinary constituents into n-words: **XP^sem = n-word**.
- N-words (but not ordinary constituents) can also be combined with *nem* (also right-adjoined to them): **NW^nem**.

Table 2 offers an overview of the distribution of NWs alone, NWs combined with *sem* or *nem*, and XPs converted into n-words by *sem*.

<table>
<tr><th colspan="2">PREVERBAL DOMAIN</th><th rowspan="2">VERB</th><th rowspan="2">POSTVERBAL DOMAIN</th></tr>
<tr><th>VP-adjoined</th><th>[Spec,VP]</th></tr>
<tr><td>{ NW* YP^snem | NW* }</td><td>YP^snem</td><td></td><td>{ YP^sem | NW }</td></tr>
<tr><td>const. neg.</td><td>const. neg. & clause neg.</td><td></td><td>const. neg.</td></tr>
<tr><td colspan="4">YP^snem: { NW^sem | NW^nem | XP^sem }
YP^sem: { NW^sem | XP^sem }</td></tr>
</table>

Table 2. The distribution of n-words

conjunction. *Sem*, in turn, developed into a minimizing particle, the negative counterpart of *is*, and into s*em… sem*…, a correlative pair of conjunctions. For further details, see É. Kiss (2011).

[5] There is one more negative particle: *ne* 'not'. Its use is constrained to imperative, subjunctive and optative sentences. In these sentences it has the same distribution and the same basic negative polarity licensing potential as *nem* 'not'. I leave the analysis of negative polarity in these sentence types (covering the distribution of all the other negative particles) and the XLE implementation of this analysis to future work.

[6] I will give the relevant examples when I present my analysis in section 4.

[7] In the distributional schemas I use the following symbols: NW = (intrinsic) n-word, XP = any constituent other than NW, YP = { XP | NW }.

- An n-word can appear without a right-adjoined negative particle postverbally, see NW in the rightmost column of the table, naturally in the presence of an appropriate preverbal negative polarity licensor, see (3).
- Even in such a (postverbal) configuration, it can be combined with *sem*, right-adjoined to it, see the NW^sem disjunct of YP^sem. (Postverbally, *nem* cannot right-adjoin to it.) For instance, in (3) we could have *senki-t se(m)* '#nobody-ACC also_not'. This version would be more emphatic, given the semantics of *sem*. Thus, in this case the contribution of *sem* is adding emphasis in the sense of 'not even'.
- This latter case and all the other cases are instances of what I loosely and informally call overt constituent negation, in which the negative particle (*sem* or *nem*) right-adjoins to the target constituent,[8] see Table 2, for the distributional facts of right-adjunction. The main empirical generalization here is that the negative particle in these configurations does not license the occurrence of other n-words in the sentence.
- *Sem* is also capable of converting an ordinary (non-n-word) into an n-word, see the second disjunct of YP^sem. This constituent is the "negative (i.e. negatable) counterpart" of YP^*is* ('YP^also').[9]
- In both the VP-adjoined position and the [Spec,VP] position, an intrinsic n-word can be combined with either *sem* or *nem* in such a way that the particle is right-adjoined to it; see the first two disjuncts of YP^snem.
- In both the VP-adjoined position and the [Spec,VP] position, a non-n-word can be converted into a ("derived") n-word by right-adjoining *sem* to it. *Nem* cannot be used in this role.
- When an intrinsic (or derived) n-word appears in [Spec,VP] both constituent and predicate negation takes place: it licenses additional n-words postverbally, also see Footnote 3. In this case, several n-words can be licensed in VP-adjoined positions; however, they must not be combined with *sem* or *nem*.
- When an intrinsic (or derived) n-word appears in a VP-adjoined position, only constituent negation takes place, and these negative polarity items licensed by *sem* or *nem* take scope over other operators to their right.
- In the VP-adjoined domain a *sem/nem*-negated intrinsic n-word can be preceded by one or more other (strictly non-negated) intrinsic n-words: NW* YPsnem$_{[NW]}$.

[8] Recall from section 2 that in ordinary constituent negation (including the negation of the universal quantifier) *nem* 'not' is employed, and it left-adjoins to the target constituent.

[9] Also see Footnote 4.

4 An LFG-XLE analysis

Let me start this section with the discussion and analysis of (inherent or derived) n-words in the postverbal domain. Consider the examples in (5) (next page).

- As these examples and (3) illustrate, n-words can occur postverbally iff they are licensed by a negative particle.
- *Sem* can turn an ordinary constituent into a (derived) n-word, compare (5a) and (5b).
- Only *sem* can right-adjoin to an intrinsic n-word in this domain, and *nem* cannot be used: (5c).[10]
- An intrinsic n-word can be used on its own (without being combined with a right-adjoined negative particle): (5d).

The sentences in (3) and (5) are ambiguous. *János* 'John.NOM' can be interpreted as (i) the focus or (ii) the topic of the sentence. In my approach the negative particle *nem* is a non-projecting word capable of occupying head-adjoined and phrasal positions. In Laczkó (2014b), in my analysis of (i) I assume that *János* 'John.NOM' occupies the regular [Spec,VP] focus position and the negative particle is left-adjoined to V^0, and in the case of (ii) I assume that *János* 'John.NOM' is in a topic position, and the negative particle is in [Spec,VP].[11] See the schematic structural representation in (1) in section 2.

(5)

János	**(nem)*	*lát*	*meg*
John.NOM	not	sees	VM

'John doesn't catch sight of

(a) *egy* *lány-t.*
a girl-ACC

(b) *egy* *lány-t* *sem.*[12]
a girl-ACC also_not

(c) *senki-t* *sem / *nem.*
#nobody-ACC also_not not

(d) *senki-t.*
#nobody-ACC

(a) a girl.' [+specific]
(b) a girl, either.' [–specific]
(c) anybody at all.'
(d) anybody.'

[10] In the preverbal domain *nem* is also usable.

[11] This assumption is strongly supported by the fact that the VM appears postverbally.

[12] É. Kiss (2015) points out that constructions like (5a) and (5b) are radically different. If a [–specific] indefinite noun phrase occurs postverbally in the scope of a negative particle, it must be combined with a right-adjoined *sem:* (5b). Otherwise it will be interpreted as a [+specific] indefinite noun phrase: (5a).

Nem in [Spec,VP] and *nem* in the V^0-adjoined position manifest the default, basic configurations for the licensing of n-words.[13] The simplest case of this is when an intrinsic n-word occurs postverbally on its own. Recall that in Laczkó (2014b) I assume that the negative particle in all its five major uses, whether involved in predicate negation or constituent negation, has its own PRED feature and it has the ADJUNCT function. When the negative particle is involved in predicate negation (in [Spec,VP] or in a V^0-left-adjoined position), it is the entire f-structure of the clause that it is an adjunct of, while in the case of constituent negation it is an adjunct of the negated constituent (XPneg in my XLE representation). The crucial question from this perspective is how we can encode the n-word licensing potential of the negative particle in the relevant cases (and the lack of this potential in the rest of the cases).

This question needs to be posited in the larger context of treating negation phenomena in the ParGram community, which has been (and has remained) an unsettled issue from the perspective of uniformity since 2006.[14] Below are the most important aspects of this issue that are immediately relevant for us here.

- There are languages in which negation is encoded by a particle, an independent word (e.g. English, Polish and Hungarian). In others, a bound morpheme, a negative suffix is used (e.g. in Turkish). In certain others, both strategies are employed (e.g. in Wolof and Indonesian). On the basis of these morphological properties, the following intuitive solution suggested itself on the ParGram line: (i) if the negative particle is an independent word, it can be assumed that it has a PRED feature and it functions as a special negative adjunct (ii) if it is a bound morpheme, then it is naturally analyzed as an element without a PRED feature that contributes the NEG+ feature. It needs to be pointed out right away that LFG's basic assumptions also naturally accommodate the opposite view: (i) a free morpheme only contributing a feature (value), (ii) a bound morpheme encoding a PRED feature. I think it was primarily due to this principled flexibility of the LFG architecture that ParGram grammars went in radically different directions in the treatment of negation phenomena. This whole issue was even more complexly challenging in

[13] Below I will show that, as I briefly mentioned in section 3, there is an alternative strategy available for n-word licensing: it is possible to combine an intrinsic n-word with a right-adjoined *nem* or *sem*, or a non-n-word with *sem* in the [Spec,VP] position, and this also provides an appropriate postverbal domain for n-words, i.e. in addition to constituent negation, it will also have the n-word licensing predicate negation effect. Moreover, this configuration also licenses VP-adjoined n-words; however, in this case they cannot combine with *sem* or *nem*.

[14] For detailed discussions, see Rákosi (2013) and Laczkó (2015).

the case of languages which employ both the free and the bound morpheme strategies. See the points below.

- In the English and Hungarian ParGram grammars the negative particles are analyzed as special negative adjuncts with their own PRED feature, see the discussion of Laczkó (2014b) in section 2. Interestingly, the Polish ParGram grammar (in its 2014 version) employed the NEG+ implementational option.[15]
- The Turkish ParGram grammar, because of the affixal nature of the relevant element, assumes that it has no PRED feature, and it only contributes the NEG+ feature.
- Although Wolof has both strategies, the current Wolof ParGram grammar uniformly applies the NEG+ analysis.
- By contrast, while Indonesian, too, makes use of both strategies, the Indonesian ParGram grammar has uniformly implemented the neg-adjunct analysis.
- In addition to the neg-adjunct and NEG+ devices, there is a third alternative: the negative specification of polarity: POL = negative. For instance, the English ParGram grammar uses this for the analysis of the following construction type: *I had no time.* The particle *no* has its own PRED feature, it is treated as a quantifier and it encodes the negative value for the POL(arity) feature.

In this general ParGram context, I augment my XLE analysis of constituent and predicate negation in Laczkó (2014b) along the following lines, in order to capture n-word phenomena as well.

- The encoding of the relevant domain for licensing n-words is a syntactic issue in Hungarian that needs to be modelled in c-structure and f-structure (from the perspective of both parsing and generation).
- I keep the neg-adjunct treatment of the negative particle. The basic generalizations are as follows.
  - In all the five basic uses analyzed in Laczkó (2014b), it has a constituent negating function. When it is left-adjoined to a non-verbal constituent (i) any constituent in [Spec,VP] or (ii) a universal quantifier in $[XP\ VP]_{VP,}$ ordinary constituent negation takes place: it is an adjunct of the given constituent, it negates it, but for obvious reasons it cannot scope out of the constituent; therefore, it cannot have a scope-taking, n-word licensing function.
  - When it left-adjoins to the verbal head (V^0) or when it occupies the [Spec,VP] position,[16] it has the n-word licensing potential.
- In section 2 I pointed out that in Laczkó (2014b) I distinguish a third type of predicate negation: VP negation, when the negative particle left-adjoins

[15] For a modified analysis, see Przepiórkowski & Patejuk (2015).
[16] See the relevant configurations in (1).

to a whole VP. Note, however, that the particle in this use is not an n-word licensor. Compare (6) with (5b), for instance. (In (6) a verbal particle, i.e. a VM, occupies the [Spec,VP] position.)

(6) *János* *nem* *meg* *lát* *valaki-t* */ *senki-t,*
John.NOM not VM sees somebody-ACC / #nobody-ACC

hanem *fel* *hív* *valaki-t.*
but VM calls somebody-ACC

intended meaning: 'It is not the case that John catches sight of somebody, instead, he calls up somebody.'

The above facts have the following consequences.

- It would not be appropriate to encode the n-word licensing effect of the negative particle by including the following specification in its lexical form (in one way or another): when it is an adjunct of any projection of a verb, it automatically contributes a feature[17] to the f-structure of the clause that licenses n-words. Instead, this has to be structurally encoded in the V^0-adjoined and [Spec,VP] cases. Notice that in all the three configurations the negative particle is an adjunct of the entire clause, but it is not capable of licensing n-words when it left-adjoins to a VP.
- It is important to note that in the case of (non-n-word-licensing) VP negation the negative particle only has scope over the VP (it cannot scope to the left, so topics are not in its scope: they have wide scope). In an important sense then this is an instance of constituent negation (VP-negation). The scope relationships can be straightforwardly captured by the f-precedence device.
- The previous point also provides an additional argument against analyzing this negation type by dint of the NEG+ feature, because such a feature cannot naturally be involved in f-precedence relationships.
- In addition to the previous two points, there is a further fact that lends considerable support to the neg-adjunct analysis: VP-negation and predicate negation of the [Spec,VP] type can be combined. Consider the following sentence.

(7) *János* *nem* *NEM* *lát* *meg* *senki-t,*
John.NOM not NOT sees VM #nobody-ACC

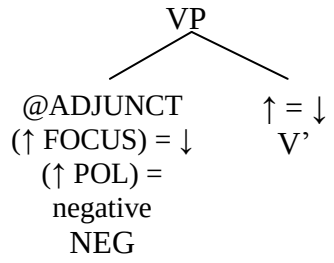
hanem *NEM* *hív* *fel* *senki-t.*
but NOT calls VM #nobody-ACC

'It is not the case that John doesn't catch sight of anybody; instead, he does not call anybody up.'

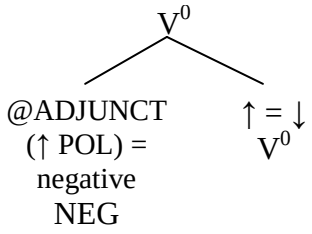
[17] I discuss the nature of this feature below.

- The most natural way of modelling the sensitivity of n-words to the presence of a domain licensed by the negative particle is by making their occurrence dependent on a feature introduced by the negative particle. Recall that in the ParGram inventory currently there are three devices used for handling negation facts: neg-adjunct, NEG+ and POL = negative. As I pointed out above in a different context, it would not be an appropriate solution to constrain the appearance of n-words to the presence of neg-adjunct in the f-structure of the clause, because it is there in the case of VP negation as well, but VP negation does not license n-words. I think that the most natural feature here is POL = negative. This truly and even mnemonically expresses the essence of this phenomenon: n-words are negative polarity items.[18] I also think that the NEG+ device is most felicitously used for affixal negation, as in the Turkish ParGram grammar. I would find it counterintuitive to assume that a bound morpheme, attached to the verb stem, encodes a neg-adjunct.
- On the basis of the above considerations, in this augmented approach I maintain my treatment of the negative particle in Laczkó (2014b) as regards its lexical representation, see (2) in section 2.
- I assume that its n-word licensing potential must be associated with two of its possible syntactic occurrences: in the V^0-adjoined position and in [Spec,VP]. See the representations in (8a) and (8b) below.

(8) a.



b.



The first two annotations in [Spec,VP] and the first annotation in the V^0-adjoined position are the same as in my earlier analysis in Laczkó (2014b),[19] and I have simply added the (↑ POL) = negative annotation, which n-words are to be represented as being sensitive to. In other words, the appropriate environment for n-words is c-structure-annotationally encoded. Naturally, it

[18] POL = negative could be treated in XLE either as an ordinary feature or as a CHECK feature. In my analysis I use the former solution because it more straightforwardly captures the fact that n-words and their licensors, the negative particles, are in various (semantic) scope relations, for details, see É. Kiss (2015), for instance. CHECK features, by contrast, simply ensure syntactic well-formedness (by checking certain constellations of constituents).

[19] The @ADJUNCT template introduces the following annotation: ↓ ∈ (↑ ADJUNCT).

also has to be encoded that the following (inherent or derived) n-words can occur in the postverbal domain: XP^sem, NW^sem and NW, see (5b), (5c) and (5d), respectively. In the current version of HunGram I have implemented the first two cases by the following two phrase structure rules.

(9) Vbar → V
YPsem: @YP-GF[20]
(↑ POL) =c negative.[21]

This encodes the fact that one of the possible sisters of V below V' is a special constituent with the YPsem label. Such labelling is rather standard in the XLE tradition: it even mnemonically signals the nature of this constituent: an ordinary constituent is combined with the right-adjoined *sem* particle. Such specific c-structure labels contribute to parsing and generating efficiency. @YP-GF is the usual template for the range of grammatical functions this constituent can have, and crucially the constraining equation restricts the occurrence of this constituent to the presence, in the f-structure of the clause, of the POL = negative feature-value pair.

(10) YPsem → YP
SEM: @ADJUNCT.

This rule encodes the fact that any constituent can be combined with a right-adjoined element of category SEM with an adjunct function. The lexical form of *sem* is given in (11).[22]

(11) *sem* SEM (↑ PRED) = 'sem'
(↑ ADJUNCT-TYPE) = neg.

[20] In my rules I use the YP label, because the relevant range of categories is DP, ADVP and PP, and XP is reserved for a larger, more general range of categories in other syntactic positions.

[21] In this analysis I only cover negative polarity in finite clauses and leave developing an account of negation in non-finite (infinitival) clauses to future work. One of the differences will be that the polarity annotation for the YPsem constituent in the Vbar of an infinitival clause will contain an extended path: (↑ XCOMP* POL) =c negative. This will encode the fact that a negative particle in an appropriate licensing position in the finite matrix clause has the entire infinitival clause in its negative polarity scope. I thank one of my reviewers for asking a question that prompted me to point this out. Consider the following example.

(i) *Kati* *nem* *akar* *meg* *lát-ni* *senki-t.*
Kate.NOM not wants VM see-INF #nobody-ACC
'Kate doesn't want to catch sight of anybody.'

Here the n-word object of the infinitive is licensed by the negative particle in the [Spec,VP] position of the finite matrix clause.

[22] Recall that the other variant of this particle, *se*, behaves in the same way in all possible respects; therefore, it has exactly the same lexical form.

It is worthwhile comparing the rules and representations of ordinary constituent negation with the *nem* particle in my analysis in Laczkó (2014b) and those of this special constituent negation with *sem*. In section 2 I gave the lexical form for *nem* in (2), which I repeat here as (12) for convenience.

(12) *nem* NEG (↑ PRED) = 'nem'
(↑ ADJUNCT-TYPE) = neg.

And my phrase structure rule for constituent negation in Laczkó (2014b) is given in (13) below.

(13) XPneg → NEG: @ADJUNCT;
XP.[23]

The formal parallels between (10) and (13), on the one hand, and between (11) and (12), on the other hand, are straightforward. In addition, they are also similar semantically: they are used to express constituent negation in these configurations.[24]

As (5d) illustrates, an intrinsic n-word can also occur in the postverbal negative polarity domain on its own (without the "support" of *sem*). I have implemented this by dint of the following annotated phrase structure rule.

(14) Vbar → V
XP: @XP-GF
{ (↓ POL-TYPE) ~= negative
|(↓ POL-TYPE) =c negative (↑ POL) =c negative }.

In the second line, the @XP-GF template is the usual grammatical function specification for postverbal constituents. In the current system, n-words are specified as belonging to the quant PRON-TYPE, and their polarity is negatively specified: (↑ POL-TYPE) = negative. On the basis of this, the disjunction in the third and fourth lines encodes the following: the XP is not an n-word ***or*** if it is an n-word, the f-structure of the clause must contain the POL = negative feature-value pair. For this analysis to work, I use the following V^0-left-adjunction, i.e. Vneg, rule.

(15) Vneg → NEG: @ADJUNCT
(↑ FOCUS PRED FN) ~= nem
(↑ FOCUS POL-TYPE) ~= negative
(↑ POL) = negative;
V.

@ADJUNCT is the usual adjunct template. The negative constraint in the second line ensures that the [Spec,VP] and this NEG position cannot be

[23] XP here ranges over all the major non-verbal categories: DP, PP, AP and ADVP.
[24] For further details of their similarities and differences, see below.

simultaneously filled by the negative particle.[25] The constraint in the third line encodes the fact that in this configuration [Spec,VP] cannot be occupied by an n-word.[26] And the equation in the fourth line introduces the negative polarity domain.

Consider the examples in (17) and (18).

(16) *János* *sem* *lát* *meg* *senki-t.*
John.nom also_not sees VM #nobody-ACC
'John does not catch sight of anybody, either.' or:
'Neither / Not even John catches sight of anybody.'

(17) *Senki* *senki-vel* *nem/sem* *lát* *meg* *senki-t.*
#nobody.NOM #nobody-with not/also_not sees VM #nobody-ACC
'Nobody catches sight of anybody with anybody (at all).'

(18) *Senki* *senki-vel* *nem/sem* *KATI-T* *látja*
#nobody.NOM #nobody-with not/also_not KATE-ACC sees

meg *(*senki-nél).*
VM #nobody-at

'Nobody catches sight of KATE with anybody at anybody's place.'

They illustrate the following empirical generalizations I made in section 3.

- *Sem* can turn an ordinary constituent into a (derived) n-word by right-adjoining to it, and when this combination occupies the [Spec,VP] it functions as a negative polarity licensor, see (16).
- When an intrinsic n-word in [Spec,VP] is combined with either *nem* or *sem*, also right-adjoined to it, the same negative polarity licensing takes place. In this case, left-VP-adjoined n-words are also licensed by this NW^sem/nem; however, in such positions they must not be combined with *sem* or *nem*, see (17)
- In the VP-adjoined domain a *sem/nem*-negated intrinsic n-word can be preceded by one or more other (strictly non-negated) intrinsic n-words: NW* YPsnem$_{[NW]}$, see (18)

My rules for the treatment of (16) and (17) are as follows.

[25] Given that I use the neg-adjunct treatment of the negative particle, this makes it very convenient and straightforward for me to encode this constraint, because I can (negatively) indicate the PRED value without argument structure (i.e. PRED FN) of the particle in the focus position. This would be much more complicated in a NEG+ approach.

[26] The reason for this is that, as I will show next, when an n-word occupies the focus position, *sem* or *nem* must be right-adjoined to it, and this complex will encode the negative polarity licensing (POL=negative) feature-value pair as well, and in this case V^0-left-adjunction is blocked.

(19) VP → YPsnem: (↑ POL) = negative;
Vbar.

Just like in the postverbal domain, where I use YPsem, in the preverbal domain, too, I use a special c-structure category: YPsnem. The major difference between them is that the postverbal variant can only contain *sem*, while the preverbal one can also contain *nem* if it is right-adjoined to an intrinsic n-word.[27] The only annotation associated with YPsnem is the marking of the negative polarity domain. All the other aspects are encoded in the c-structure rule for YPsnem in (20).

(20) YPsnem → { YP: @YP-GF
(↓ POL-TYPE) ~= negative
{ (↑ FOCUS) = ↓ | (↑ FOCUS) }
SEM
| YP: @YP-GF
(↓ POL-TYPE) =c negative
{ (↑ FOCUS) = ↓ | (↑ FOCUS) }
{ SEM
| NEG} }.

In the first main disjunct I model the combination of an ordinary constituent and *sem*. The particle can only be *sem* (*nem* is excluded). The (↓ POL-TYPE) ~= negative constraint makes sure that only ordinary (i.e. non-n-word) constituents are involved. The { (↑ FOCUS) = ↓ | (↑ FOCUS) } disjunction handles the distribution of YPsnem. It can only occur in the preverbal domain in two positions: (a) in [Spec,VP], see the first disjunct: it will be the focused constituent; (b) in $[XP,VP]_{VP}$, see the second disjunct: it requires the presence of focus elsewhere (i.e. in [Spec,VP]).

In the second main disjunct the (↓ POL-TYPE) =c negative equation constrains this configuration to n-words. The function of the { (↑ FOCUS) = ↓ | (↑ FOCUS) } disjunct is the same as that of the similar disjunct in the first main disjunct. Finally, the { SEM | NEG } disjunction encodes the fact that either *sem* or *nem* can right-adjoin here.

[27] The label *snem* is meant to indicate this: *sem* or *nem*. Let me also point out that it is one of the most controversial empirical and theory-sensitive issues whether it can be assumed that *senki sem* '#nobody also_not' and *senki nem* '#nobody not' have an isomorphic structure or not (i.e. whether *nem* is also really right-adjoined to the n-word). In my generalizations and analysis here I assume this isomorphism without justification, and I will argue for this assumption when I compare my LFG account with mainstream GB/MP approaches in future work. I will present functional, distributional and prosodic arguments.

The relevant c-structure rule for (18) is as follows.

(22) VPquantneg → { YPsnem: (↑ FOCUS)
(↑ FOCUS PRED FN) ~= nem
(↑ FOCUS POL-TYPE) ~= negative
| YP+: (↓ POL-TYPE) =c negative;
YPsnem: (↓ PRON-TYPE)
(↑ FOCUS)
(↑ FOCUS PRED FN) ~= nem
(↑ FOCUS POL-TYPE) ~= negative }
VP.

The first disjunct in the disjunction handles the case in which there is only a single derived (i.e. non-pronominal) or non-derived n-word in the adjoined position. (↑ FOCUS) encodes the fact that YPsnem can be VP-adjoined if there is a focused constituent in [Spec,VP]. (↑ FOCUS PRED FN) ~= nem expresses the fact that this constituent is not the negative particle. (↑ FOCUS POL-TYPE) ~= negative means that this YPsnem cannot co-occur with YPsnem in [Spec,VP]. In the second disjunct YP+ with its (↓ POL-TYPE) =c negative annotation encodes the fact that optionally the single obligatory YPsnem can be preceded by one or more n-words. This captures the generalization that the occurrence of n-words in a VP-adjoined position is conditional on the presence of a single YPsnem phrase, i.e. it is in this way that YPsnem licenses an n-word in a pre-VP position.[28]

Now consider the example in (23) and its c-structure representation in Figure 1.

(23) *Soha* *senki* *senki* *mellett* *sem* *lát* *meg*
#never #nobody.NOM #nobody beside also_not sees VM

senki-t *sehol* *sem* *János-sal* *sem.*
#nobody-ACC #nowhere also_not John-with also_not

'Nobody catches sight of anybody anywhere beside anybody ever also without John.'

[28] This is the current implemented encoding of n-word licensing in this configuration, which seems to be the simplest solution, and the most efficient one from the perspective of both parsing and generation. Notice, however, that in this case the n-words preceding YPsnem are not licensed by the (↑ POL) = negative feature; instead, the presence of a right-adjacent, negated n-word is the licensor. In future work, when I compare my LFG account with mainstream GB/MP approaches, I will return to this issue by also taking other possible LFG-XLE solutions into consideration and assessing their strengths and weaknesses. At this stage let me only point out that YPsnem in the VP-adjoined position is not a negative polarity licensor for the VP domain. This fact may yield independent motivation for treating this case differently.

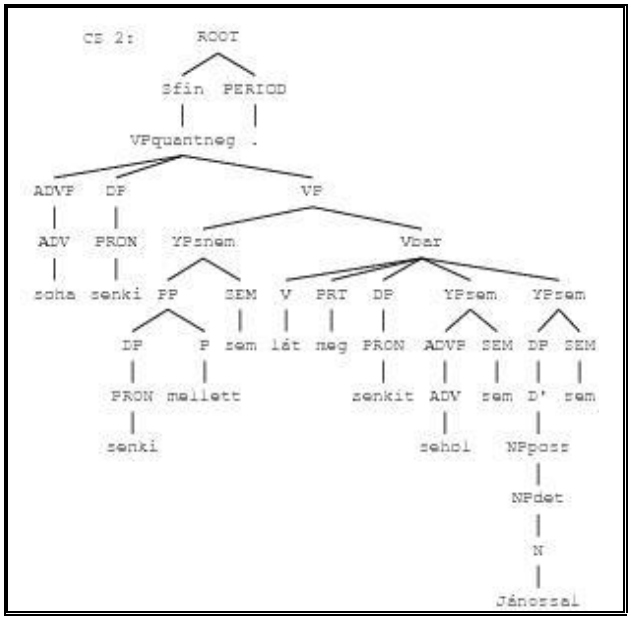


Figure 1. The c-structure of (23)

In this sentence an YPsnem constituent occupies the [Spec,VP] position, and it licenses the two VP-left-adjoined n-words as well as the postverbal negative polarity items: an n-word on its own (DP), an n-word combined with *sem* (the first YP) and an ordinary constituent combined with *sem*.

Finally, let me show that I have extended this implemented analysis of negative polarity to the two suppletive forms of the copula *van* 'be'. As is well-known, in certain functions[29] the indicative, present tense, 3SG and 3PL forms are: *nincs* 'not.be.PRES.3SG', *nincsenek* 'not.be.PRES.3PL', *sincs* 'also_not.be.PRES.3SG', *sincsenek* 'also_not.be.PRES.3PL'. Consider the following examples.

(24) *János / JÁNOS nincs senki-nél.*
John.NOM John.NOM not.be.3SG #nobody-at
'John/JOHN isn't at anybody's place.'

(25) *JÁNOS sincs senki-nél.*
JOHN.NOM also_not.be.3SG #nobody-at
'JOHN isn't at anybody's place, either.'

(26) *Senki SENKI-VEL nincs / sincs senki-nél.*
#nobody.NOM #nobody-with (also_)not.be.3SG #nobody-at
'Nobody is at anybody's place with anybody (either).'

As (24) shows, if a constituent precedes *nincs*, the sentence is ambiguous, and the constituent can be interpreted as either the topic or the focus of the

[29] In the existential, locative and possessive uses of the copula.

sentence. (25) demonstrates the fact that the constituent preceding *sincs* must be interpreted as the focus. (26) illustrates the fact that an n-word can be combined with either *nincs* or *sincs*, cf. its combinability with either *nem* or *sem*. Here I can only show the c-structures of (25) and (26) in Figure 2. In the latter, I show the *nincs* version.

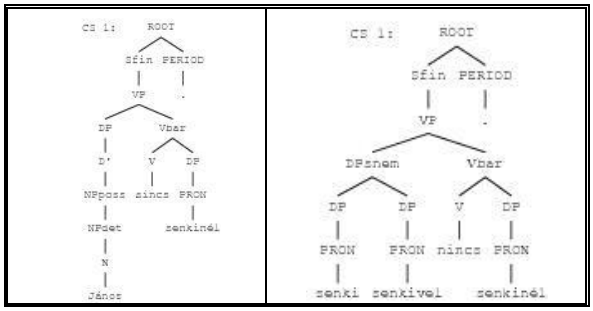


Figure 2. The c-structures of (25) and (26)

The crucial aspects of my analysis are as follows. I use the following lexical form for *nincs*.

(27) *nincs* V (↑ PRED) = 'nincs < (↑ SUBJ) (↑ OBL) >'
(↑ POL) = negative
(↑ NEG) = +
{ ~(↑ FOCUS)
| (↑ FOCUS)
(↑ FOCUS PRED FN) ~= nem }.

In the PRED value I give the actual (singular) form of the copula: *nincs*. The argument structure is that for the locative use of the copula. This word itself encodes negative polarity. Notice that this phenomenon is a strong argument from Hungarian for the dual neg-adjunct and NEG+ approach that I am proposing here for the following reason. Typically, negation in Hungarian is marked by a syntactic atom, a negative particle, which in certain configurations also introduces a negative polarity domain (but not always even in the case of predicate negation). However, these suppletive forms merge the usual copula features, predicate negation and the negative polarity feature. This can be taken to be an extreme instance of the affixal encoding of negation and negative polarity. For this reason, in the lexical forms of *nincs* and *sincs* I use the NEG+ feature. It would be highly counterintuitive (although it would, of course, be technically possible) to handle this along the neg-adjunct lines, by using the usual neg-adjunct annotations. *Sincs* has the same lexical form, except that it requires the [Spec,VP] position to be filled obligatorily by a focussed element: an n-word or an ordinary constituent; thus, instead of the disjunction in (27) it only has the (↑ FOCUS) annotation.

I also need to modify my YPsnem rule, because in these copula constructions the YPsnem constituent must not contain *nem/sem*, because

negation is encoded by the special negative forms of the copula *nincs/sincs*. This can be captured by adding the following disjunct, which itself contains two disjuncts, to the YPsnem rule.

(28) YPsnem → { YP: @YP-GF
(↓ POL-TYPE) =c negative
{ (↑ PRED FN) =c nincs
| (↑ PRED FN) =c sincs }
(↑ POL) =c negative
(↑ FOCUS) = ↓
| YP: @YP-GF
(↓ POL-TYPE) ~= negative
(↑ PRED FN) =c sincs
(↑ POL) =c negative
(↑ FOCUS) = ↓ }

The peculiarity of this disjunct is that the rule does not contain SEM or NEM: it simply rewrites YPsnem as YP for intrinsic n-words and for ordinary constituents, see the values of the (↓ POL-TYPE) attribute in the two disjuncts. Both disjuncts are constrained to a special negative polarity environment, see (↑ POL) =c negative, in which the predicate is *nincs* or *sincs* in the case of intrinsic n-words and *sincs* in the case of ordinary constituents, which is captured by the (↑ PRED FN) =c nincs/sincs equations.

5 Concluding remarks

In this paper, I have modified and augmented my LFG-XLE analysis of negation proposed in Laczkó (2014b) by (i) developing an account of the special uses of negative particles (ii) capturing their interaction with negative polarity items (iii) presenting a formal treatment of the two forms of the two suppletive negative variants of the copula.

In order to ensure parsing and generating efficiency, I have made use of the standard XLE devices: special syntactic categories: NEG and SEM, and specifically labelled phrasal projections: YPsnem and YPsem.

I have argued for using all the three modes of treating negation phenomena in the analysis of Hungarian.

In the spirit of Forst et al. (2010) and Laczkó & Rákosi (2011) I use the non-projecting categories PRT and NEG in both head-adjunction and phrasal configurations. This is different from Toivonen's (2001) proposal. She assumes that certain categories in Swedish have projecting and non-projecting variants. The non-projecting versions are head-adjoined to the verb and the projecting versions have the regular phrasal behaviour. Note that this approach could also be straightforwardly accommodated in my analysis: head-adjoined NEG vs. phrasal NEGP. However, I am not aware of any phrasal projection property of the negative particle; that is why I treat it

uniformly as a non-projecting word. Moreover, technically it would also be possible to do without the non-projecting treatment. Instead of assuming that the negative particle is left-head-adjoined to the verb when the focus position is filled by a constituent: NEG^V^0, one could assume that NEGP left-adjoins to V'.[30]

In general, the special functional categories NEM and SEM, and the specifically labelled phrasal nodes YPsnem and YPsem could also be dispensed with. It would be possible to assume that negative particles are adverbs and they project ADVPs, and these (special) ADVPs occupy the positions my non-projecting NEGs and SEMs occupy. Naturally, such an approach would conform to standard X-bar-syntactic assumptions and conventions to a greater extent. The cost would be that a more complex system of constraining equations and CHECK features would be needed to prevent overgeneration from the perspective of both parsing and generation. In future work I will set out to explore the behaviour and a possible (generalized) treatment of a range of "small words" in Hungarian including preverbs, *csak* 'only', *is* 'also', *volna* (the marker of irrealis mood), *-e* (the yes-no question marker), *nem* 'not', *ne* 'not' in imperative, subjunctive and optative sentences, *se(m)* 'also_not', and I will address such general aspects of possible alternative approaches. One of the most likely conclusions of my investigation will be that LFG's architecture and assumptions make it possible to capture generalizations about such complex phenomena in an explicit and principled way based on the trade-off between c-structure and f-structure representations.

Acknowledgements

I thank the participants of the LFG15 conference for useful remarks. I am also grateful to my anonymous reviewers, because their very helpful comments prompted me to extend and sharpen my analysis as well as to improve certain presentational aspects of the paper. I am solely responsible for all remaining errors.

This research was supported by the Hungarian Scientific Research Fund (OTKA), grant No. 111918.

References

É. Kiss, Katalin. 1992. Az egyszerű mondat szerkezete [The structure of the simple sentence]. In: Kiefer, Ferenc. ed. *Strukturális magyar nyelvtan 1. Mondattan.* [Structural Hungarian grammar 1. Syntax]. Budapest: Akadémiai Kiadó, 79-177.

[30] For instance, in her GB framework, É. Kiss (1992) has a V'-adjunction analysis and in É. Kiss (1994) she assumes V^0-adjunction.

É. Kiss, Katalin. 1994. Sentence structure and word order. In: Kiefer, Ferenc & É. Kiss, Katalin. eds. *The Syntactic Structure of Hungarian*. New York: Academic Press, 1-90.

É. Kiss, Katalin. 2008. Tagadás vagy egyeztetés? A *senki, semmi* típusú névmások szórendi helye, jelentése és hangsúlyozása [Negation or concord? The word order, interpretation and prosody of se-pronouns] *Magyar Nyelv* 104 (2): 129-143.

É. Kiss, Katalin. 2011. A *sem* szinkrón és diakrón szerepéről [On the syncronic and diacronic roles of *sem*]. In: Kádár, Edit & Szilágyi, M. Sándor. eds. *Szinkronikus nyelvleírás és diakrónia* [The synchronic description of languages and diachrony]. Kolozsvár: Erdélyi Múzeum Egyesület, 95-109.

É. Kiss, Katalin. 2015. Negation in Hungarian. In: Tamm, Anne, Miestamo Matti & Wagner-Nagy, Beáta. eds. *Negation in Uralic Languages*. *Typological Studies in Language* 108. Amsterdam: John Benjamins Publishing Company, 219-238.

Forst, Martin; King, Tracy H. & Laczkó, Tibor. 2010. Particle verbs in computational LFGs: Issues from English, German, and Hungarian. In: Butt, Miriam & King, Tracy H. eds. *Proceedings of the LFG '10 Conference*. On-line publication. Stanford: CSLI Publications, 228-248.

Laczkó, Tibor & Rákosi, György. 2008-2014. *HunGram. An XLE Implementation.* University of Debrecen.

Laczkó, Tibor & Rákosi, György. 2011. On particularly predicative particles in Hungarian. In: Butt, Miriam & King, Tracy H. eds. *Proceedings of the LFG '11Conference*. Stanford: CSLI Publications, 299-319.

Laczkó, Tibor. 2014a. Essentials of an LFG analysis of Hungarian finite sentences. In: Butt, Miriam & King, Tracy H. (eds.) *Proceedings of the LFG14Conference*. Stanford: CSLI Publications, 325-345.

Laczkó, Tibor. 2014b. Outlines of an LFG-XLE account of negation in Hungarian sentences. In: Butt, Miriam & King, Tracy H. (eds.) *Proceedings of the LFG14Conference*. Stanford: CSLI Publications, 304-324.

Laczkó, Tibor. 2015. *On an LFG-XLE treatment of negation in Hungarian.* Presentation at the ParGram Meeting in Warsaw, 4 February, 2015. *ParGram Wiki: Pargram Topics:* http://typo.uni-konstanz.de/redmine/projects/pargram/wiki/Negation.

Przepiórkowski, Adam & Patejuk, Agnieszka. 2015. *Two representations of negation in LFG: Evidence from Polish.* Talk presented at LFG'15, Waseda University, Tokyo, July 20, 2015.

Rákosi, György. 2013. Negation. Presentation at the 2013 Debrecen ParGram Meeting. 23 July, 2013. *ParGram Wiki: Pargram Topics:* http://typo.uni-konstanz.de/redmine/projects/pargram/wiki/Negation.

Toivonen, Ida. 2001. *The Phrase Structure of Non-Projecting Words*. Ph.D. dissertation, Stanford University.

ON OPERATORS IN THE PREVERBAL DOMAIN OF HUNGARIAN FINITE SENTENCES

Tibor Laczkó
University of Debrecen

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King
(Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

In this paper, I concentrate on the distribution and interaction of operators (question phrases, focused constituents, universal quantifiers and the negative particle) in the preverbal domain of Hungarian finite clauses. I considerably modify the LFG-XLE analysis I presented in Laczkó (2014a). On the one hand, I complement some aspects of my earlier analysis. On the other hand, I enlarge the empirical coverage greatly by developing an account of all the construction types investigated by Mycock (2010).

1 Introduction

In a series of recent papers, I presented the essential ingredients of the most comprehensive LFG analysis of Hungarian finite clauses to date (designed to be XLE-implementable). In Laczkó (2014a) I developed the crucial aspects of an LFG-XLE syntax of the preverbal portion of finite sentences. In Laczkó (2014b) I outlined an analysis of various constituents in the [Spec,VP] position: foci, question phrases, and several types of verbal modifiers (VMs). In Laczkó (2014c) I concentrated on sentence and constituent negation. In this paper, I will augment the programmatic approach presented in Laczkó (2014a,b) by making the analysis of the basic construction types more complete and more constrained, on the one hand, and by incorporating the analysis of further constructions, on the other hand. I will cover the phenomena analyzed by Mycock (2010), who investigated the marking of scope relations from both syntactic and prosodic perspectives.

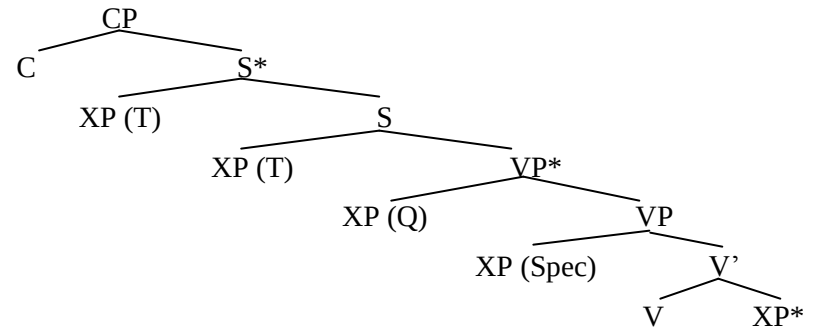
The structure of the paper is as follows. In section 2, I summarize the relevant points of my earlier analysis in Laczkó (2014a). In section 3, I discuss Mycock (2010) in a detailed fashion. In section 4, I modify and augment my earlier analysis in order to give an LFG-XLE account of all the constructions studied by Mycock. In section 5, I make some concluding remarks.

2 On Laczkó (2014a)

In Laczkó (2014a), I posit the generalized sentence structure shown in (1) on next page.[1] Table 1 gives an overview of the disjunctive annotations associated with the topic field, the quantifier zone, and the [Spec,VP] position schematically represented in (1). Given that in this paper I focus on constituents in the last two positions, I will only make very brief remarks on the programmatic treatment of elements in the topic field, which will need to be further developed (e.g. to capture the narrow scope of contrastive topics).

[1] XP* is the customary Kleene star notation (any number of XPs and possibly none), while S* and VP* indicate the iterativity of $[\text{XP S}]_{\text{S}}$ and $[\text{XP VP}]_{\text{VP}}$ adjunction.

(1)



T: **{ (c-)topic | sent.adv. }**	**Q:** **{ quantifier | WH }**	**Spec:** **{ focus | WH | VM }**
{ (↑ GF) = ↓ {↓ ∈ (↑ TOPIC) | ↓ ∈ (↑ CONTR-TOPIC)} | ↓ ∈ (↑ ADJUNCT) (↓ ADV-TYPE) =c SENT }	(↑ GF) = ↓ { (↓ CHECK _QP) =c + | (↑ CHECK _VM-INTER) =c + (↓ CHECK _QP-INTER) =c + (↓ SPECIFIC) =c + }	{ (↑ GF) = ↓ (↑ FOCUS) = ↓ | (↑ GF) = ↓ (↓ CHECK _VM-INTER) =c + ((↑ CHECK _VM-INTER) = +) | { (↑ GF) = ↓ | ↑ = ↓ } (↓ CHECK _VM) =c + }

Table 1. Basic functional annotations in the left periphery in Laczkó (2014a)

In the topic field there are three basic annotational possibilities for a constituent. (A) It has either an argument function or an adjunct function (represented as GF in a generalized way: (↑GF) = ↓), and it has (i) a topic or (ii) a contrastive topic discourse function. (B) It has an adjunct function if it is a sentence adverb: (↓ ADV-TYPE) =c SENT.

The most important aspects of the annotations in the Q ("quantifier") column of Table 1 are as follows.[2]

- A constituent in this field bears a grammatical function, and it is either a quantifier or a question phrase. This is encoded by the disjunction.
- In the two disjuncts, I use the XLE-style CHECK featural device. Its essence is that these CHECK features come in pairs: there is a defining equation and it has a constraining equation counterpart. These CHECK feature pairs can ensure that two elements will occur together in a

[2] Parts of this description have been taken from Laczkó (2014a). In the annotational disjunctions the disjuncts are in complementary distribution. This is formally encoded in XLE by negatively specifying each disjunct for the (positive) features of all other disjuncts. For the sake of legibility, these negative specifications are omitted throughout the paper.

particular configuration,[3] or that a particular element will occur in a designated position. It is this latter property that I utilize here.

- In the first disjunct, the constraining CHECK feature equation requires a constituent containing an element that is (inherently) specified as a quantifier.[4] The defining CHECK feature equation counterpart is included in the lexical entries of the quantifier elements involved, see the generalized lexical form representation in (2).

(2) L (quantifier) …
(CHECK _QP (GF*↑)) = +

- The second disjunct regulates the occurrence of additional question phrases in multiple constituent questions. The combination of the (↑ CHECK _VM-INTER) =c + and the (↓ CHECK _QP-INTER) =c + constraining equations guarantees that this position can be occupied by an interrogative expression (second equation) iff the [Spec,VP] position is already occupied by another interrogative expression (first equation). Question words are assumed to have the generalized lexical form shown in (3). The annotations encode the following properties respectively.

a) These elements are interrogative pronouns.
b) They occur in constituent questions.
c) They occur in sentences that do not contain a focused constituent.[5]
d) They are constrained to occur in the [Spec,VP] or the (VP-adjoined) quantifier positions.

(3) L (wh-word) …
(↑ PRON-TYPE) = interrogative
(STMT-TYPE (GF* ↑)) = wh-interrogative
~(FOCUS (GF* ↑))
{ (CHECK _VM-INTER (GF* ↑)) = +
| (CHECK _QP-INTER (GF* ↑)) = + **}**

And now I turn to the annotations I associate with the [Spec,VP] position.

- The three main disjuncts encode the complementary distribution of focused constituents, question phrases and VMs, respectively.

[3] For an example of this, see Laczkó & Rákosi's (2011) treatment of Hungarian particle verb constructions, in which the simplex verb and the particle are marked by corresponding CHECK features in their respective lexical forms.
[4] _QP is mnemonic of this category.
[5] This captures the fact that, on the one hand, question phrases and ordinary focused constituents are in complementary distribution, aspiring to the same [Spec,VP] position, and, on the other hand, even when one or several of them do not occur in [Spec,VP] that position has to be occupied by another question expression (and not a focused constituent). This latter generalization will be modified when the last construction type is analyzed in this paper.

- The first disjunct is straightforward.[6]
- In the second disjunct, the first (constraining) CHECK feature equation requires the presence of a question phrase in this position. Its defining counterpart is given in the lexical forms of question words, see (3).
- In the second disjunct, the second, optional, defining CHECK feature equation serves as the licensor of the occurrence of question phrases in the quantifier field.[7] When it is not present in the structure, no question phrase can occur in the quantifier position. When it is present, it requires the presence of one or more question phrases. From the perspective of question phrases in the quantifier position, they can only occur there if the [Spec,VP] position is filled by a question phrase.
- The third disjunct handles VMs. The defining counterpart of its constraining CHECK feature equation is included in the lexical forms of the elements that can occupy this position in neutral sentences (in non-focused sentences and non-constituent-question sentences). The functional head annotation (↑ = ↓) in the disjunction is for particles, while the (↑ GF) = ↓ annotation is for all the other types of VMs.

3 On Mycock (2010)

I discuss Mycock's paper in a detailed fashion here for the following reasons. (i) This work reports the results of very important experimental research (based on elicited spoken data) exploring the syntax-prosody interface with respect to encoding prominence in Hungarian. (ii) It covers a wide range of phenomena, and posits its account in an LFG framework. (iii) Some aspects of Mycock's syntactic view that I do not agree with are shared by several other researchers, so when I discuss these details I can argue in a generalized fashion against similar proposals.

Let me present Mycock's (2010) overview of the intonation patterns she attested in her experimental research and her assumptions about the syntax of these constructions.[8] Consider her table and her examples, whose numbers are also included in the table (2010: 285). Below I number these examples as (M21)…(M39).

[6] Although I subscribe to the very strong recent view in LFG that discourse functions are to be uniformly represented in i-structure, for the sake of simplicity of exposition here I apply the classical LFG representation of TOPIC and FOCUS in f-structure.

[7] Its constraining counterpart is associated with the VP-adjoined position.

[8] In this paper I can only concentrate on the syntactic aspects of Mycock's (2010) approach. Her claims about the syntax-prosody interface with respect to scope marking are as follows. (i) Either syntax or prosody can mark scope (by dint of word order and stress, respectively). (ii) They typically align. (iii) When they do not align, prosody wins out, i.e. prosody determines scope. My view of this issue is greatly different, but, due to limitations of space, I have to discuss this elsewhere.

Predicate					
Operator Field					
QP	FOCUS		VERB	POSTVERBAL FIELD	Example
---------	focus		*verb*	*VM DO LOC*	(21)
---------	NEG + focus		*verb*	*VM*	(23)
---------	single Q-phrase		*verb*	*VM DO*	(31)
∀	----------		*VM + verb*	*SUBJ*	(24)
∀	----------		*NEG + verb*	*DO*	(29)
NEG + ∀	----------		*verb*	*VM SUBJ*	(28)
∀	focus		*verb*	*VM*	(25)
---------	**Q1**	Q final	*verb*	*VM DO*	(33)
---------	**Q1 Q2**	Q final	*verb*	*VM*	(35)
---------	**Q1**	NEG + focus	*verb*	*VM*	(39a)
---------	Q1	NEG + focus	*verb*	*VM*	(39b)

Table 2. General patterns of intonation

A dashed line indicates that no constituent occupies the relevant syntactic position. The point of prosodic prominence (a sharply falling pitch accent H+L at the left edge of the first phonological word) is represented by shading; the low plateau which follows it is indicated by italics; any high (H) monotone preceding the H+L accent is indicated by bold.

(M21) [János]$_{\text{TOPIC}}$ [**Anná-nak**]$_{\text{FOCUS}}$ mutatta be Mari-t
John.NOM *Anna*-DAT *introduced* VM *Mary*-ACC
a mozi-ban.
the cinema-INE
'John introduced Mary to ANNA at the cinema.'[9]

(M23) János **nem**=[Mari-t]$_{\text{FOCUS}}$ hívta fel.
John.NOM NEG=*Mary*-ACC *called* VM
[lit.] 'John called not MARY.'
('John called someone other than Mary.')

[9] I have simplified Mycock's glossing of the past tense verbs (by not indicating the encoding of agreement and (in)definiteness). Throughout the paper, in the examples I will follow Mycock's convention of italicizing the words in the glosses.

(M31) [János]$_{\text{TOPIC}}$ [**ki-nek**]$_{\text{FOCUS}}$ mutatta be Mari-t?
John.NOM *who*-DAT *introduced* VM *Mary*-ACC
'Who did John introduce Mary to?'

(M24) **Mindenki-t** fel-hívott János.
everyone-ACC VM-*called* *John*.NOM
'For every *x*, *x* = person, John called *x*.'

(M29) Mindenki **nem**=dicsérte Anná-t.
everyone.NOM NEG=*praised* *Anna*-ACC
'Not everyone praised Anna.'

(M28) **Nem**=mindenki-t hívott fel János.
NEG=*everyone*-ACC *called* VM *John*.NOM
[lit.] 'John called not everyone.'
('Not everyone was called by John.')

(M25) **Mindenki-t** [János]$_{\text{FOCUS}}$ hívott fel.
everyone-ACC *John*.NOM *called* VM
'For every *x*, *x* = person, JOHN called *x*.'

(M33) [Ki]$_{\text{FOCUS}}$ [**ki-nek**]$_{\text{FOCUS}}$ mutatta be Mari-t?
who.NOM *who*-DAT *introduced* VM *Mary*-ACC
'Who introduced Mary to who?'

(M35) [Ki]$_{\text{FOCUS}}$ [ki-t]$_{\text{FOCUS}}$ [**ki-nek**]$_{\text{FOCUS}}$ mutatott be?
who.NOM *who*-ACC *who*-DAT *introduced* VM
'Who introduced who to who?'

(M39) [János]$_{\text{TOPIC}}$ [ki-nek]$_{\text{FOCUS}}$ nem=[Mari-t]$_{\text{FOCUS}}$ mutatta be?[10]
John.NOM *who*-DAT NEG=*Mary*-ACC *introduced* VM
[lit.] 'Who did John introduce not MARY to?'
('Who did John introduce someone other than MARY to?')

In Table 3 I juxtapose Mycock's syntactic analysis of the relevant construction types with my syntactic analysis to be developed here.[11] As the top of Table 3 shows, Mycock subscribes to the widely assumed basic sentence articulation in Hungarian, where phrase-structurally the verb heads a VP, focus is in [Spec,VP] and the postverbal field is dominated by V'. The actual structural treatment of quantifiers is not stated (whether they are VP-adjoined or they are sisters of VP). By contrast, I assume the modified articulation also shown at the top of Table 3. In my view, too, focus is in [Spec,VP]. However, I assume that all VM types are also in [Spec,VP] in

[10] (39a) and (39b) in Table 2 only differ in their prosody; that is why just one example is given in (M39).
[11] Naturally, the prosodic aspects of the constructions under investigation are the same.

complementary distribution with focus.[12] In addition, in my approach, constituents in the operator field are left-adjoined to VP.

Mycock (2010)				**Laczkó (this paper)**			
Predicate				Predication			
Operator Field			VERB	Operator Field	Predicate		
QP	FOCUS			QP	Spec,VP	VERB	Ex.
---------	focus		verb	---------	focus	verb	21
---------	NEG+focus		verb	---------	NEG+focus	verb	23
---------	single Q-phrase		verb	---------	single Q-phrase	verb	31
∀	----------		VM+verb	∀	VM	verb	24
∀	----------		NEG+verb	-------- (!)	NEG	verb	29
NEG+∀	----------		verb	---------	NEG+∀	verb	28
∀	focus		verb	∀	focus	verb	25
---------	**Q1**	Q final	verb	**Q1**	Q final	verb	33
---------	**Q1 Q2**	Q final	verb	**Q1 Q2**	Q final	verb	35
---------	**Q1**	NEG+focus	verb	**Q1**	NEG+focus	verb	39a
---------	Q1	NEG+focus	verb	Q1	NEG+focus	verb	39b

Table 3. Comparison of Mycock's (2010) and my syntactic analysis

Elsewhere I will argue in a detailed fashion for distinguishing two functional notions in sentence articulation in Hungarian: predication vs. predicate. My predication corresponds to Mycock's predicate, and both of them correspond to the comment component of the customary topic-comment primary division. My predicate is the core VP, and the VP-adjoined operators are in my operator field.

Let me now discuss the similarities and differences between our syntactic analyses of the relevant construction types.

(21), (23), (31). The two analyses are the same – the constituents receiving prosodic prominence are in the syntactically designated [Spec,VP] focus position.

(24). Mycock: the quantifier is in QP; the [Spec,VP] position, which is reserved for focused constituents in her approach, is empty; the VM morphologically combines with the verb (i.e. both elements are under V^0). Laczkó: the quantifier is in QP here, too; the [Spec,VP] position is also a

[12] Contra Mycock (2010), Payne & Chisarik (2000) and Gazdik (2012), for instance.

standard position for VMs, so the VM occupies this position; and the simplex verb is under V^0.[13]

(29). Mycock: the quantifier is in QP; the [Spec,VP] position, which is reserved for focused constituents in her approach, is empty; and the negative particle even syntactically procliticizes to the verb under V^0. Laczkó: the quantifier is ***not*** in QP here: it is in a (contrastive) topic position preceding the QP,[14] hence the "--------- (!)" representation in Table 3; the [Spec,VP] position is also a standard position for the negative particle, so NEG occupies this position, and the simplex verb is under V^0.[15]

(28). Mycock: the negated universal quantifier is in its regular ("cartographic") QP position; [Spec,VP] is empty; and the verb is in V^0. Laczkó: the negated universal quantifier is in [Spec,VP], just like any ordinary negated constituent, which must be focused as a rule;[16] and the verb is in V^0.

(25). The two analyses are the same: the universal quantifier, receiving the H+L prosodic prominence, is in QP; the focused constituent is in its usual [Spec,VP] position, but this time without its usual H+L accent; and the verb is in V^0.

(33), (35). Mycock: all the question phrases make up one cluster that occupies the [Spec,VP] focus position. Laczkó: it is always the final (immediately preverbal) question phrase that occupies the [Spec,VP] position; and all the non-final question phrases are in the operator field, in left-VP-adjoined positions.

(39a,b). The fundamental difference between the two approaches is the same as in the case of the previous construction type. Mycock: the (non-immediately-preverbal) question phrase and the negated (non-interrogative) focus make up a cluster, which is the focused unit in [Spec,VP]. Although she does not mention this, this treatment is highly problematic for Mycock's approach for an analysis-internal reason. One of her fundamental assumptions is that interrogative foci (one question phrase or a cluster of question phrases) and a(n always single) non-interrogative focused

[13] For my detailed argumentation against assuming that VMs and verbs make up a morphological unit and for assuming that VMs are in [Spec,VP], see Laczkó (2014b).

[14] The crucial evidence for the contrastive topichood of the quantifier in this construction is that it can intermingle with other (ordinary or contrastive) topics.

[15] My main argument in Laczkó (2014c) for positing the negative particle in [Spec,VP] is that its complementarity with other constituents also targeting [Spec,VP] is naturally and directly captured. Let me add here that the heavy stress on the particle attested by Mycock's (2010) pitchtrack for this example lends further support to this view. For the details of my approach see my analysis of example (6) in section 4.

[16] As I point out in Laczkó (2014c), a negated universal quantifier can occupy the QP position iff the [Spec,VP] position is filled by a non-negated focused constituent.

constituent are in complementary distribution in [Spec,VP]. This is clearly violated by (39). Laczkó: only the negated (non-interrogative) focus is in [Spec,VP]; and the (non-immediately-preverbal) question phrase is in a left-VP-adjoined position (in the operator field).[17]

4 Augmenting Laczkó's (2014a) analysis

Below I analyze the constructions in the same order as they were presented and discussed in the previous section.

(M21) [János]$_{\text{TOPIC}}$ [**Anná-nak**]$_{\text{FOCUS}}$ mutatta be Mari-t
John.NOM *Anna*-DAT *introduced* VM *Mary*-ACC
a mozi-ban.
the cinema-INE
'John introduced Mary to ANNA at the cinema.'

Other than my remarks on my earlier account, I have nothing to add about the treatment of constituents in the topic field; thus, the analysis of the topic constituent *János* 'John' is as usual. In this example there is no constituent in the operator field. The oblique argument, *Annának* 'to Anna' is the focus in the [Spec,VP] position. In Table 4, I show the relevant annotations in my previous account and those in my new analysis.

Laczkó (2014a)	**Laczkó (this paper)**
[Spec,VP]	**[Spec,VP]**
(↑ GF) = ↓ (↑ FOCUS) = ↓	(↑ GF) = ↓ (↑ VM-FOCUS) = ↓ { (↓ VM-FOCUS-TYPE) = exh [↗=↙, ρ: erad] | (↓ VM-FOCUS-TYPE) = id [↗=↙, ρ: level] | (↓ VM-FOCUS-TYPE) = pres { [↗=↙, ρ: level] | [↗=↙, ρ: erad] } }

Table 4. Functional annotations for focus in [Spec,VP]

In the new analysis, too, (↑ GF) = ↓ is the same standard generalized grammatical function annotation as in my previous analysis. The (↑ FOCUS) = ↓ annotation in the previous analysis is radically augmented here. The reason for this is that in the previous, programmatic approach I only modelled one focus type in this single designated [Spec,VP] position, the generally assumed exhaustive type. In work in progress (still concentrating on the preverbal domain in Hungarian sentences) I also treat a

[17] The difference between the two variants in (39a) and (39b) is that different preverbal constituents receive prosodic prominence, the H+L pitch accent.

construction type in which a focused constituent occurs in the operator field, preceding a question phrase in [Spec,VP]. Therefore, the two foci (the "standard" one in [Spec,VP] and this other one in this special construction) need to be distinguished. My solution is that I label the standard focus as VM-FOCUS[18] and all other occurrences of foci (in either the preverbal or the postverbal domain) simply as FOCUS. Partially motivated by Kálmán et al. (1984), Kálmán (1985, 2001), Kálmán et al. (1986) and Gazdik (2012), in this work in progress I distinguish three types of focus that constituents can be associated with: ordinary exhaustive focus, presentational focus and identificational focus; the third one roughly corresponds to Kálmán's (2001) and Gazdik's (2012) "hocus". And there is also a special, additional type, often called verum focus, whose function is to verify the truth of a statement.[19] In the representation of my new analysis in Table 4 I use the [↗=↙, ρ: level/erad] notation[20] as an informal, short-hand representation for a complete set of prosodic annotations (along the lines of Mycock 2006 and Dalrymple & Mycock 2011). The labels "level" and "erad" stand for the prosodic properties of the constituent in this position in level-prosody and eradicating-stress sentence types, respectively.[21]

(M23) János **nem**=[Mari-t]$_{\text{FOCUS}}$ hívta fel.
John.NOM NEG=*Mary*-ACC *called* VM
[lit.] 'John called not MARY.'
('John called someone other than Mary.')

In my implemented LFG-XLE analysis of constituent negation in Laczkó (2014c) I use the following XLE-style c-structure rules and functional annotations. I disjunctively add the XPneg phrasal category to the usual XP category in [Spec,VP], and it receives the customary (↑GF)=↓ (↑FOCUS)=↓ annotations.[22] My phrase structure rule for the XPneg itself is as follows.

(4) XPneg → NEG: ↓ ∈ (↑ ADJUNCT);
XP.[23]

[18] I use the "VM" prefix in the function label to indicate that the focused constituent in this [Spec,VP] position competes with VMs. In addition, VMs, too, can be focused there.

[19] In the case of this focus type (which is also often called VP-focus), too, the H+L accent falls on the constituent in [Spec,VP] if that position is filled. If it is not filled, the verb is stressed.

[20] The ρ symbol stands for c-structure—p-structure linking.

[21] In Mycock's (2010) analysis it has the characteristic H+L accent; in Mycock's (2006) representation it has the ↙βTONE=fall annotation.

[22] In my current analysis the negated focused constituent is also associated with the [↗=↙, ρ: erad] prosodic annotation.

[23] XP is the functional head of XPneg. Recall that it is a convention in XLE that the ↑ = ↓ functional head annotations are not indicated in the phrase structure, and the

NEG is used to implement the idea that the negative particle is a special non-projecting functional word which can be adjoined to X^0 and XP categories (in this case it adjoins to an XP).[24] NEG is analyzed as an adjunct. I assume the following lexical form for the negative particle.

(5) *nem* NEG (↑ PRED) = 'nem'
(↑ ADJUNCT-TYPE) = neg.

Its category is NEG. This word is a special adjunct expressing negation, which is encoded by the (↑ ADJUNCT-TYPE) = neg annotation.

(M31) [János]$_{\text{TOPIC}}$ [**ki-nek**]$_{\text{FOCUS}}$ mutatta be Mari-t?
John.NOM *who*-DAT *introduced* VM *Mary*-ACC
'Who did John introduce Mary to?'

Laczkó (2014a)	**Laczkó (this paper)**
[Spec,VP]	**[Spec,VP]**
(↑ GF) = ↓ (↓ CHECK _VM-INTER) =c + ((↑ CHECK _VM-INTER) = +)	[↗=↙, ρ: erad] (↑ GF) = ↓ (↓ CHECK _VM-INTER) =c + ((↑ CHECK _VM-INTER) = +)

Table 5. Functional annotations for a question phrase in [Spec,VP]

The only difference between my previous account and my current analysis is that in the latter I also indicate the (exhaustive focus type) eradicating prosody of the constituent (and sentence). When there is a single question phrase in the sentence, occupying the [Spec,VP] position, it only has the first, obligatory CHECK feature annotation. The second, optional CHECK feature is needed for the treatment of multiple constituent questions, as discussed above in connection with Table 1.

(M24) **Mindenki-t** fel-hívott János.
everyone-ACC VM-*called* *John*.NOM
'For every *x*, *x* = person, John called *x*.'

Recall that this is Mycock's (2010) example with her representation of the particle+verb combination as a single word; and also recall that in my approach the particle is an independent word occupying the [Spec,VP] position, just like other VM constituents.[25] For the details of my analysis of

system automatically associates this annotation with nodes lacking other (grammatical) functional annotations.

[24] In future work I will argue, in the spirit of Toivonen (2001) but with partially different assumptions, that there are several non-projecting categories in Hungarian, and this group includes the negative particle.

[25] The universal quantifier *mindenki* 'everyone' is in the operator field in both Mycock's and my analysis.

particle verb constructions, see Laczkó (2013), and for the details of my general treatment of VMs, see Laczkó (2014b).

Laczkó (2014a)	**Laczkó (this paper)**
[Spec,VP]	**[Spec,VP]**
{ (↑ GF) = ↓ | ↑ = ↓ } (↓ CHECK _VM) =c +	[↗=↙, ρ: level] { (↑ GF) = ↓ | ↑ = ↓ } (↓ CHECK _VM) =c +

Table 6. Functional annotations for VMs in [Spec,VP]

Recall that in my previous analysis the (constraining) CHECK feature guarantees that only elements lexically specified as VMs can occupy this position in a "neutral" sentence. The ↑ = ↓ functional head annotation is for particles and the (↑ GF) = ↓ annotation is for all the other VM types. These annotations are retained in my new analysis as well. However, here I also indicate the characteristic prosodic properties of VMs under normal (i.e. level prosodic) circumstances.

As regards the treatment of the universal quantifier, compare my two accounts in Table 7.

Laczkó (2014a)	**Laczkó (this paper)**
$[XP,VP]_{VP}$	**$[XP,VP]_{VP}$**
(↑ GF) =↓ (↓ CHECK _QP) =c +	[↗=↙, ρ: erad] (↑ GF) =↓ (↓ CHECK _QP) =c +

Table 7. Functional annotations for universal quantifiers in $[XP,VP]_{VP}$

In my previous analysis, the programmatic annotation is very simple. In my new analysis, I schematically indicate the prosodic properties of the quantifier: [↗=↙, ρ: erad]. This encoding expresses that the universal quantifier gets the H+L pitch accent in this neutral construction type with VMs. In the discussion of the analysis of (M25), I will repeat the empirical generalization that even when a universal quantifier is followed by a focus, it is the former that receives the H+L accent**.**

(M29) Mindenki **nem**=dicsérte Anná-t.
everyone.NOM NEG=*praised* *Anna*-ACC
'Not everyone praised Anna.'

Recall from my discussion of Table 3, comparing the crucial aspects of Mycock's (2010) and my view of the construction types under investigation, that Mycock assumes that, on the one hand, the universal quantifier in (M29) is in the operator field, and, on the other hand, the negative particle procliticizes to the verb. Mycock's representation of these assumptions in the table shows that for her this procliticization (indicated as *nem=dicsérte* in

(M29)) is not only prosodic but also syntactic. By contrast, above I argued that, on the one hand, the universal quantifier has all the properties of a contrastive topic, and, thus, it occupies a position in the topic field, and, on the other hand, the negative particle is in the [Spec,VP] position. As regards the treatment of the universal quantifier, it has to be associated with the functional annotations for contrastive topics.

The essence of my LFG-XLE analysis of this type of predicate (or clausal) negation in Laczkó (2014c) is as follows. My main argument for positing that the negative particle is in [Spec,VP] is its complementarity with the other elements competing for this position: focused constituents, question phrases and VMs.[26] This complementarity is most straightforward in the case of sentences containing VMs. Consider (6), for instance, in which there is a preverb as a VM, occurring postverbally.

(6) Péter nem hívta fel a barátjá-t.
Peter.NOM NEG *called up the friend.his*-ACC
'Peter didn't call up his friend.'

I assume that, in addition to the other three types of elements targeting the [Spec,VP] position, NEG has to be included in a fourth disjunct with the following XLE style annotations.

[Spec,VP]
↓ ∈ (↑ ADJUNCT) (↑ VM-FOCUS) = ↓ (↓ VM-FOCUS-TYPE) = neg [↗=↙, ρ: erad]

Table 8. Functional annotations for NEG

As this table shows, I assume that NEG in [Spec,VP] has the FOCUS function.[27] My motivation for this is twofold. On the one hand, the negative particle's prosody is identical to that of an ordinary focused constituent.[28] On the other hand, in the current version of our HunGram grammar, the complementarity, in this construction type, of the negative particle and the VM can be captured (implemented) in a straightforward way: the general rule is that the VM targets the [Spec,VP] position provided that it is not occupied by a focused element, and *nem* is one such element.

[26] For detailed discussion, including the summary of some analyses that do not assume that NEG also competes for the [Spec,VP] position, see Laczkó (2014c).

[27] Naturally, this view makes it necessary to augment the generally assumed inventory of focus types. I leave the discussion of this issue to another forum.

[28] Consider the following quote from Mycock (2010) in this connection. "In terms of its prosody, an utterance involving preverbal negation shares key features with other non-neutral sentences, such as those which include a syntactically focused constituent […]. The negative particle *nem* bears Kálmán et al.'s (1986) 'eradicating stress', i.e. a sharp fall in pitch followed by a low plateau" (2010: 276).

(M28) **Nem**=mindenki-t hívott fel János.
NEG=*everyone*-ACC *called* VM *John*.NOM
[lit.] 'John called not everyone.'
('Not everyone was called by John.')

Recall that Mycock assumes that in the case of (M28) the negated universal quantifier is in its regular QP position, the focus position is empty, and, despite this fact, the VM occurs postverbally. By contrast, I argue in Laczkó (2014c) that a negated universal quantifier can occupy its canonical QP position if and only if the [Spec,VP] position is filled by a non-negated focused constituent.[29,30] From this it follows that in my approach *nem mindenkit* 'not everyone.ACC' in (M28) is analyzed in exactly the same way as *nem Marit* 'not Mary.ACC' in (M23).

(M25) **Mindenki-t** [János]$_{\text{FOCUS}}$ hív-ott fel.
everyone-ACC *John*.NOM *call*-PAST.3SG VM
'For every *x*, *x* = person, JOHN called *x*.'

Recall that this construction is analyzed in the same way syntactically by Mycock (2010) and me. As has been pointed out several times above, it is a special prosodic property of this construction type that the universal quantifier "steals" the H+L pitch accent from the (exhaustive) focus. This can be captured in my system in the following way. I need to ensure that the two designated constituents "see each other" from and in their respective positions. The representational strategy is the same as in my treatment of multiple questions: I use CHECK feature pairs, see the relevant part of the discussion of Table 1 above. The key idea here is that the CHECK feature in the quantifier position ensuring (constraining) that only (universal) quantifiers can occur in that position,[31] (↓ CHECK _QP) =c + is supplemented with an optional defining CHECK feature *with an up-arrow*:

[29] This immediately explains the postverbal occurrence of the VM.

[30] Note that in Mycock's analysis the negated universal quantifier receives the H+L pitch accent, because it is a universal quantifier in its canonical position, while in my analysis it receives this accent because it is a negated constituent in the focus position, that is why it is associated with the [↗=↙, ρ: erad] notation. It would be interesting to explore experimentally, by using minimal pairs, whether a non-negated universal quantifier and its negated counterpart exhibit exactly the same prosodic behaviour, and whether the negated quantifier has exactly the same prosodic properties in the following two configurations: NEG+∀ verb and NEG+∀ focus verb. If there was some noticeable difference, that would lend additional support to my analysis. However, if there was no discernible contrast, that would not necessarily support Mycock's view.

[31] In the disjunct of annotations for quantifiers, as opposed to question phrases.

(↑ CHECK _QP) = +, whose constraining counterpart is associated with the exhaustive focus in [Spec,VP]: (↑ CHECK _QP) =c +.[32]

$[XP,VP]_{VP}$		$[XP,VP]_{VP}$
[↗=↙, ρ: erad] (↑ GF) =↓ (↓ CHECK _QP) =c +	→	[↗=↙, ρ: erad] (↑ GF) =↓ (↓ CHECK _QP) =c + ((↑ CHECK _QP) = +)

Table 9. Functional annotations for universal quantifiers in $[XP,VP]_{VP}$ to capture the prosody of co-occurrence with exhaustive focus

And the combination of this constraining CHECK feature disjunctively with the regular eradicating stress annotation associated with exhaustive focus will have the following effect. In the unmarked case the focused constituent will have eradicating stress, but there will be no prosodic annotation, i.e. there will be no eradicating stress associated with the focus if there is a universal quantifier in $[XP,VP]_{VP}$.

(↑ GF) = ↓ (↑ VM-FOCUS) = ↓ (↓ VM-FOCUS-TYPE) = exh [↗=↙, ρ: erad]	→	(↑ GF) = ↓ (↑ VM-FOCUS) = ↓ (↓ VM-FOCUS-TYPE) = exh { [↗=↙, ρ: erad] | ~[↗=↙, ρ: erad] (↑ CHECK _QP) =c + }

Table 10. Modification of prosodic annotations for the exhaustive focus preceded by a universal quantifier

The disjunction part of the annotations encodes the following information. The first disjunct is the prosodic annotation I have used so far. The second disjunct encodes the fact that exhaustive focus has no eradicating stress: ~[↗=↙, ρ: erad] if there is a universal quantifier in $[XP,VP]_{VP}$: (↑ CHECK _QP) =c +. In this case the quantifier will receive eradicating stress, see Table 9.

Mycock (2010) does not exemplify and analyze the "NEG+∀ focus verb" construction type. Let me show how I can treat it in my approach. Consider the example in (7), and compare it with (M28) and (M25).

[32] In the case of multiple questions, it is the immediately preverbal question phrase in [Spec,VP] that receives a similar pair of CHECK features: (↓ CHECK _VM-INTER) =c + and ((↑ CHECK _VM-INTER) = +). The second, optional one licenses additional question phrases in the quantifier position.

(7) Nem mindenki-t János hív-ott fel.
NEG *everyone*-ACC *John*.NOM *call*-PAST.3SG VM
[lit.] 'JOHN called not everyone.'
('It doesn't hold for everyone that it was John who called them.')

János 'John.NOM' is a focused constituent and it is preceded by a universal quantifier (which happens to be negated); therefore, it must be analyzed in the same way as *János* 'John.NOM' in (M25), where it is preceded by a non-negated universal quantifier. The annotational apparatus needs to be augmented in the quantifier position to capture the fact that a universal quantifier can be negated in its canonical position iff there is a focused constituent in [Spec,VP].[33] Consider the annotations from Table 7 above and the modification I propose here.

Table 7	**Modified version**
$[XP,VP]_{VP}$	**$[XP,VP]_{VP}$**
[↗=↙, ρ: erad] (↑ GF) =↓ (↓ CHECK _QP) =c +	[↗=↙, ρ: erad] (↑ GF) =↓ (↓ CHECK _QP) =c + ((↓ POL) =c neg (↑ VM-FOCUS-TYPE) =c exh)

Table 11. Modification of the functional annotations for universal quantifiers in $[XP,VP]_{VP}$

The modification is very simple and straightforward: it takes the form of combining two constraining equations, and making this combination optional (the fourth and fifth annotations in the column on the right hand side). This combination encodes the following dual condition: (i) the polarity of the quantifier must be negative; AND (ii) there must be a (VM-FOCUS) discourse function in the sentence and the type of this focus must be exhaustive, which is tantamount to saying that [Spec,VP] must be filled by the standard focus type.

(M33) $[Ki]_{FOCUS}$ $[\textbf{ki-nek}]_{FOCUS}$ mutatta be Mari-t?
who.NOM *who*-DAT *introduced* VM *Mary*-ACC
'Who introduced Mary to who?'

Recall that Mycock (2010) and I analyze multiple constituent questions rather differently. The fundamental difference is that Mycock assumes that all question phrases (forming a cluster) occupy the [Spec,VP] focus position,[34] while I posit that it is solely the final question phrase that occurs in [Spec,VP], and all the other (non-final) ones are in VP-adjoined quantifier

[33] It does not matter whether the focused constituent is negated or it is not negated.
[34] (M33) is Mycock's example, showing the essence of her analysis: both question phrases are marked as being focused.

positions.[35] Consider the annotations for the treatment of multiple questions in my previous analysis taken from Table 1.[36]

[XP VP]$_{VP}$: **{ quantifier | WH }**	**[Spec,VP]:** **{ focus | WH | VM }**
(↑ GF) = ↓ (↑ CHECK _VM-INTER) =c + (↓ CHECK _QP-INTER) =c + (↓ SPECIFIC) =c +	(↑ GF) = ↓ (↓ CHECK _VM-INTER) =c + ((↑ CHECK _VM-INTER) = +)

Table 12. Basic functional annotations for the treatment of multiple questions

In the spirit of my current analysis, the annotations for the question phrase in [Spec,VP] need to be supplemented with the customary prosodic information characteristic of exhaustive focus: [↗=↙, ρ: erad].

(M39) [János]$_{TOPIC}$ [ki-nek]$_{FOCUS}$ nem=[Mari-t]$_{FOCUS}$ mutatta be?
John.NOM *who*-DAT NEG=*Mary*-ACC *introduced* VM
[lit.] 'Who did John introduce not MARY to?'
('Who did John introduce someone other than MARY to?')

I did not analyze this special construction earlier. In the general frame of the analysis I am developing in this paper, it can be treated in the following way.

(A) Fundamentally, the special and unpredicted occurrence of the question phrase needs to be encoded in the annotations for multiple questions associated with the quantifier position.

Laczkó (Table 1)	**Laczkó (augmented version)**
[XP,VP]$_{VP}$	**[XP,VP]$_{VP}$**
(↑ GF) = ↓ (↑ CHECK _VM-INTER) =c + (↓ CHECK _QP-INTER) =c + (↓ SPECIFIC) =c +	{ (↑ GF) = ↓ (↑ CHECK _VM-INTER) =c + (↓ CHECK _QP-INTER) =c + (↓ SPECIFIC) =c + | (↑ VM-FOCUS-TYPE) =c exh (↑ VM-FOCUS POL) =c neg (↓ CHECK _QP-INTER) =c + (↑ CHECK _QP-INTER) = + (↓ SPECIFIC) =c + { [↗=↙, ρ: erad] ~(↑$_{ρ}$ VM-FOCUS [↗=↙, ρ]) = erad | ~[↗=↙, ρ: erad] (↑$_{ρ}$ VM-FOCUS [↗=↙, ρ]) = erad } }

Table 13. Modified annotations for question phrases in [XP,VP]$_{VP}$

[35] It is the immediately preverbal (final) question phrase that receives the H+L accent, as the shading indicates.
[36] For details, see the discussion of Table 1 above.

Recall that in my earlier treatment of multiple constituent questions, I use the annotations shown in the left column of Table 13 (taken from Table 1). Two constraining CHECK features ensure that a question phrase can occur in this quantifier position: (↓ CHECK _QP-INTER) =c +, if the [Spec,VP] position is occupied by another question phrase: (↑ CHECK _VM-INTER) =c +. In order to cover the special construction type in (M39), this treatment needs to be augmented by the disjunction shown in the right column of Table 13. Its first disjunct is the previous set of annotations for multiple questions (see the left column again), and the second disjunct handles the special construction. The annotational strategy is basically the same here, too. A question phrase is licensed to occur in the quantifier position: (↓ CHECK _QP-INTER) =c +, if the [Spec,VP] position is occupied by a designated constituent type. Here this designated constituent is an exhaustive focus: (↑ VM-FOCUS-TYPE) =c exh that is negated (that is, its polarity is negative): (↑ VM-FOCUS POL) =c neg. The prosodic disjunction in this second disjunction formally captures Mycock's (2010) empirical findings: either the question phrase in the quantifier position (first prosodic disjunct) or the negated exhaustive focus in [Spec,VP] receives the H+L pitch accent (i.e. eradicating stress).

(B) All this has to be coupled with a modification in the annotations associated with the exhaustive focus in [Spec,VP]. Recall that the exhaustive focus, as a rule, gets eradicating stress, except when it is preceded by a universal quantifier, in which case it is the universal quantifier that receives eradicating stress. I captured this by the modified annotations in Table 10. In Table 14, I modify those annotations to also cover the prosodic behaviour of the question phrase + negated exhaustive focus construction.

(↑ GF) = ↓ (↑ VM-FOCUS) = ↓ (↓ VM-FOCUS-TYPE) = exh { [↗=↙, ρ: erad] | (↑ CHECK _QP) =c + }	→	(↑ GF) = ↓ (↑ VM-FOCUS) = ↓ (↓ VM-FOCUS-TYPE) = exh { [↗=↙, ρ: erad] | (↑ CHECK _QP) =c + | (↑ CHECK _QP-INTER) =c + }

Table 14. Modification of prosodic annotations for the exhaustive focus preceded by a universal quantifier or a question word

Recall that in the case of the "universal quantifier + focus" construction type I formally encoded the fact that the two elements "see each other" by using an optional up-arrow defining CHECK feature associated with the universal quantifier: (↑ CHECK _QP) = +, and an obligatory constraining CHECK feature associated with the exhaustive focus, and if this feature match requirement is satisfied then the focus has no eradicating stress (and the quantifier has this stress as usual).[37] In the case of our "question phrase +

[37] See the second disjunct in the left column in Table 14. In essence, this is the way in which I capture the context-sensitivity aspect of these constructions.

negative focus" construction, I also employ an up-arrow defining CHECK feature associated with the question phrase in [Spec,VP]: (↑ CHECK _QP-INTER) = +, but this time it is obligatory, because it appears among the annotations directly linked to a negated exhaustive focus,[38] and its constraining counterpart, (↑ CHECK _QP-INTER) =c +, is included in the prosodic disjunction of annotations associated with the focus, see the right column in Table 14. The scenario is the same: the focus has no eradicating stress, or, rather, it is not specified for eradicating stress when the [Spec,VP] position is filled by a question phrase (see the third disjunct). The distribution (i.e. alternation) of eradicating stress is encoded by the annotations associated with the question phrase in the right column in Table 13.

5 Conclusion

In this paper I have considerably modified my LFG-XLE analysis of the preverbal domain of Hungarian finite sentences presented in Laczkó (2014a). I concentrated on the distribution and interaction of operators (question phrases, focused constituents, universal quantifiers and the negative particle). I complemented some aspects of my earlier analysis. I enlarged the empirical coverage to a great extent by developing an account of all the construction types investigated by Mycock (2010).

Acknowledgements

I thank the participants of the LFG15 conference for useful remarks and discussions. I am also grateful to my anonymous reviewers for their comments, which enhanced the presentational aspects of the paper considerably. Any errors that remain are solely mine.

The research reported here was supported by the SROP-4.2.2.B-15/1/KONV-2015-0001 project. The project has been supported by the European Union, co-financed by the European Social Fund.

References

É. Kiss, Katalin. 1992. Az egyszerű mondat szerkezete [The structure of the simple sentence]. In: Kiefer, Ferenc. ed. *Strukturális magyar nyelvtan 1. Mondattan.* [Structural Hungarian grammar 1. Syntax]. Budapest: Akadémiai Kiadó, 79-177.

Gazdik, Anna. 2012. Towards an LFG analysis of discourse functions in Hungarian. In: Kiefer, Ferenc & Bánréti, Zoltán. eds. *Twenty Years of Theoretical Linguistics in Budapest*. Budapest: RIL HAS & Tinta Könyvkiadó, 59-92.

[38] See the (↑ VM-FOCUS POL) =c neg and (↓ CHECK _QP-INTER) =c + constraining equations in the right column in Table 13.

Kálmán, C. György, Kálmán, László, Nádasdy, Ádám & Prószéky, Gábor. 1984. Hocus, Focus, and verb types in Hungarian infinitive constructions. In: *Groninger Arbeiten zur germanistischen Lingustík, Rijksuniveriteit Groningen* 24: 162-177.

Kálmán, László. 1985. Word order in neutral sentences. In Kenesei, István. ed. *Approaches to Hungarian. Vol. 1. Data and Descriptions*. Szeged: JATE. 13-23.

Kálmán, László. 2001. ed. *Magyar leíró nyelvtan. Mondattan 1* [Hungarian descriptive grammar. Syntax 1]. Budapest: Tinta Könyvkiadó.

Kálmán, László, Prószéky, Gábor, Nádasdy, Ádám & Kálmán, C. György. 1986. Hocus, focus, and verb types in Hungarian infinitive constructions. In: Abraham, Werner & Meij, Sjaak de. eds. *Topic, Focus and Configurationality*. Amsterdam: Benjamins, 129-42.

Laczkó, Tibor. 2013. Hungarian particle verbs revisited: Representational, derivational and implementational issues from an LFG perspective. In Butt, Miriam & King, Tracy H. eds. *The Proceedings of the LFG13 Conference*. Stanford: CSLI Publications, 377-397.

Laczkó, Tibor. 2014a. Essentials of an LFG analysis of Hungarian finite sentences. In: Butt, Miriam & King, Tracy H. (eds.) *Proceedings of the LFG14 Conference*. Stanford: CSLI Publications, 325-345.

Laczkó, Tibor. 2014b. An LFG analysis of verbal modifiers in Hungarian. In: Butt, Miriam & King, Tracy H. (eds.) *Proceedings of the LFG14 Conference*. Stanford: CSLI Publications, 346-366.

Laczkó, Tibor. 2014c. Outlines of an LFG-XLE account of negation in Hungarian sentences. In: Butt, Miriam & King, Tracy H. (eds.) *Proceedings of the LFG14 Conference*. Stanford: CSLI Publications, 304-324.

Laczkó, Tibor & Rákosi, György. 2011. On particularly predicative particles in Hungarian. In Butt, Miriam & King, Tracy H. eds. *Proceedings of the LFG '11 Conference*. Stanford: CSLI Publications, 299-319.

Mycock, Louise. 2006. *A New Typology of Wh-Questions*. Ph.D. dissertation, University of Manchester.

Mycock, Louise. 2010. Prominence in Hungarian: The prosody-syntax connection. *Transactions of the Philological Society* 108(3): 265-297.

Mycock, Louise. 2013. Discourse functions of question words. In Butt, Miriam & King, Tracy H. eds. *The Proceedings of the LFG13 Conference*. Stanford: CSLI Publications, 419-439.

Payne, John & Chisarik, Erika. 2000. Negation and focus in Hungarian: An Optimality Theory account. In: *Transactions of the Philological Society* 98: 185-230.

Toivonen, Ida. 2001. *The Phrase Structure of Non-Projecting Words*. Ph.D. dissertation, Stanford University.

DEGREES OF DEGRAMMATICALIZATION: A LEXICAL SHARING APPROACH TO THE ENGLISH POSSESSIVE

John J. Lowe
University of Oxford

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King (Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

I provide a formal account of the diachronic development of the English possessive marker, from unambiguous affix in Old English, to clitic (in at least some contexts) in Present-Day English. I show that Lexical Sharing, appropriately constrained, makes it possible to represent the degrees of degrammaticalization between full clitic and full affix shown by the English possessive at different periods, while maintaining an absolute distinction between word and morpheme, as is necessary in a lexicalist theory.

1 Introduction

While most work in LFG is strictly synchronic in its aims, some researchers have sought to apply LFG to modelling syntactic change; for example, Vincent (1999, 2001a,b), Lowe (2015a: 100–121). Recently, Börjars (2013) has undertaken a comparison of how well three syntactic theories, including LFG, account for patterns in syntactic change, in particular directionality in gramaticalization.

My aim in this paper is to contribute further to exploring the value of LFG in modelling syntactic change, by providing a detailed account of (some of) the processes underlying the degrammaticalization of the English possessive marker. What was in Old English an unambiguous genitive affix in *-s* has, gradually and over a very long period of time, developed into what many authors believe is, at least in some contexts, a clitic, i.e. a distinct item in the lexicon which fills its own node in the c-structure, rather than a sub-lexical, morphological element.[1] Following the synchronic analysis advanced in Lowe (2015c), I take the status of the possessive marker in Present-Day English (PDE) to be somewhat complex. It is largely a clitic, but some instantiations of the possessive are essentially affixal, though not prototypically so. In this paper I show that the status of the possessive marker was also complex at earlier stages of the language, and I provide a formal account of the diachronic developments that must have occurred between the Old English period and the present day. My analysis makes use of 'Constrained Lexical Sharing' (Lowe 2015c), a more constrained and architecturally integrated version of Wescoat's (e.g. 2002, 2005) Lexical Sharing (LS).

[†]I am very grateful to the audience at LFG15, in particular Mary Dalrymple, Oleg Belyaev, Andy Spencer, Miriam Butt and Adam Przepiorkowski, for helpful discussion and comments, and also to the reviewers for valuable comments on this paper. I am also grateful to the audience at the 21st International Conference on Historical Linguistics, held in Oslo in 2013, where an early version of this work was presented. Related work, pulished as Lowe (2015c), was presented at LSA 2013, SE-LFG 10, and the University of Oxford General Linguistics Seminar, and I am likewise grateful to those audiences. All errors are my own. This work was undertaken while in receipt of an Early Career Research Fellowship from the Leverhulme Trust.

[1]It is a clitic in the sense that it is prosodically deficient; in syntactic terms it is no different from any other functional word.

2 PDE possessive: clitic and affix

2.1 The data

In this section I survey the most important data relating to the synchronic status of the PDE possessive marker *'s*, discussed in more detail in Lowe (2015c). The status of the PDE possessive marker has been highly controversial in previous literature. Some authors, e.g. Zwicky (1987) and Payne (2009), argue that possessive *'s* is fundamentally an affix, albeit an 'edge affix', i.e. an affix attached not to words but to syntactic phrases. Other authors, such as Quirk et al. (1985) and Anderson (e.g. 2008), argue rather that possessive *'s* is not an affix, but a clitic. In fact, there is evidence for both analyses, hence why these two irreconcilable positions both appear in the literature. The majority of data could in principle be taken either way. Usually, possessive *'s* appears directly following the possessor head noun, in one of three entirely predictable allophonic realizations:

(1) a. *Matt's coat*. [/s/, after non-sibilant voiceless segments]
b. *The angry woman's glare*. [/z/, after non-sibilant voiced segments]
c. *The goose's feathers*. [/ɪz/, after sibilant/affricate segments]

Such data can be dealt with easily under either an affixal or clitic analysis. C-structural economy of expression might favour an affixal analysis, but that principle must be counter-balanced by the need for some degree of economy within the lexicon, which would rather favour a clitic analysis.

Evidence for the affixal status of the possessive comes from lexically arbitrary variant forms. After some nouns ending in a sibilant, we find a 'zero' realization, rather than expected /ɪz/ (see Zwicky 1987: 139–141). This is found with regular plurals (2a), certain singular nouns (2b), and some proper names (2c).

(2) a. *The ducks'* (/dʌks/) *pond*. **The ducks's* (/dʌksɪz/) *pond*.
b. *The species'* (/spiːʃiːz/) *immunity*. **The species's* (/spiːʃiːzɪz/) *immunity*.
c. *Rameses'* (/ræməsiːz/) *statue*. **Rameses's* (/ræməsiːzɪz/) *statue*.

There is no regular phonological process which could account for the zero realization of the possessive marker here in these contexts, since the standard realization after a sibilant segment is /ɪz/ (1c).[2] The only alternative is to assume that the zero possessive marker is affixal, i.e. that sequences of noun followed by zero possessive are single lexical items.

[2]See e.g. Zwicky (1987: 140). Anderson (2008) assumes a morphophonological process that blocks the appearance of the clitic possessive marker /z/ when it directly follows a word ending with an inflectional morpheme /z/. This neatly explains the zero realization with regular plurals, but requires the rather implausible analysis of the underlying morphological analysis of singular *species* as [[specie]s], of *James* as [[Jame]s], and so on; this is rightly criticized by Payne (2009).

Evidence for the clitic status of the possessive comes from the 'phrasal possessive' construction. The PDE possessive marker does not necessarily attach to the head of the possessor phrase; rather it attaches to the end of the possessor phrase which, when the head is postmodified, may be a word of potentially any category:[3]

(3) a. *The Queen of England's hat*.
b. *Someone I know's brother*.
c. *The boy opposite me's sister*.
d. *The man I live with's girlfriend*.

In this 'phrasal possessive' construction, possessive *'s* is entirely unselective regarding its host. It can appear following a word of any category that can appear final in a noun phrase, including finite verbs (3b), case-marked pronouns (3c), and prepositions (3d). While it would be theoretically possible to assume that every noun, adjective, verb, preposition, etc., has a lexically specified (affixal) possessive form, this would require massive lexical duplication for little gain. Indeed, as argued in more detail in Lowe (2015c), such an approach would undermine one of the key criteria for distinguishing morphosyntactic words from larger sequences. The only solution is to treat the possessive marker as a clitic. However, when the final word in the possessor phrase is a word that takes the zero possessive, the phrase gets zero possessive marking, even when the final word in the phrase is not the head:

(4) *The female of the species'/*species's deadliness*.

In this context, then, the phrasal possessive looks rather more affixal than in (3). We are faced with conflicting evidence: some points to a clitic, some to an affixal analysis. A number of authors, including Börjars et al. (2013), emphasize that the distinction between clitic and affix is not an absolute one, and that, for example, it may be preferable to understand the two terms in reference to idealized points on a spectrum of wordhood. On some level this is true, but while in descriptive terms it might be appropriate to say that a particular morphosyntactic unit displays ambiguous properties in relation to the word–morpheme distinction, lexicalist theories of syntax such as LFG enforce an absolute distinction between word and morpheme, such that a formal representation will necessarily require an absolute categorization, one way or the other, in any given instance. In this sense, the morphosyntactic status of elements like English possessive *'s* really matters in a theory such as LFG, in a way that it would not in a non-lexicalist approach to syntax.

Börjars et al. (2013) do not set out to provide a formal analysis of the spectrum between 'ideal' clitic and 'ideal' affix; their aim is descriptive. From a more formal

[3]Work by Denison et al. (2010) and Börjars et al. (2013) demonstrates that the phrasal possessive is avoided in spoken English corpora; it is nevertheless perfectly grammatical and the consequences of this must be taken into account in any formal treatment of the PDE possessive.

perspective, Lowe (2015c) argues that it is possible to admit both clitic and affixal analyses of the PDE possessive simultaneously, providing a formal means of appropriately capturing its complex status within a lexicalist theory. My analysis of the PDE possessive is summarized in §§2.2–2.3.

2.2 Constrained Lexical Sharing

Wescoat (e.g. 2002, 2005) proposes a theory of 'Lexical Sharing' (LS) within LFG that permits an account of phenomena which, like the PDE possessive, display properties of both single-word sequences (i.e. word + affix) and two-word sequences (i.e. word + clitic). Lowe (2015c) proposes some emendations to Wescoat's model, under the name of 'Constrained Lexical Sharing', which better integrate it with recent approaches to the LFG architecture, and which more appropriately constrain its sphere of application.[4] The main differences from Wescoat's original theory are: 1. Wescoat's l-structure is identified with the syntactic string of Kaplan (1989) and standard in more recent proposals regarding the LFG architecture; 2. Lexical Sharing is absolutely restricted to contexts where it is syntactically justified, i.e. where there is clear evidence that single lexical items pattern as two-word sequences in the c-structure.

The first emendation provides a better integration of LS with most recent approaches to the LFG architecture, such as those by Dalrymple and Mycock (2011), Asudeh (2012), Mycock and Lowe (2013), and Lowe (2015a,b).[5] What for Wescoat was a projection λ from terminal c-structure nodes to a set of word forms is identified with the inverse of the π projection from the string to c-structure. I assume an architecture that includes at least the projections in (5).

(5) String $\xrightarrow{\pi}$ C-structure $\xrightarrow{\phi}$ F-structure $\xrightarrow{\sigma}$ S-structure

The second emendation places an important constraint on Wescoat's theory; under Wescoat's proposals (e.g. Wescoat 2005: 482), there are no constraints on what sorts of sequences can be analysed under LS. So Wescoat permits LS even in the case of sequences that show no evidence of being a single lexical unit (e.g. due to the presence of morphophonological irregularities). But this would endanger the very concept of a word: there would be nothing to stop wholly unambiguous sequences of 2+ words from being treated as single elements at l-structure / the string. Under 'Constrained Lexical Sharing' only sequences for which there is positive evidence for a lexical analysis (e.g. morphophonological idiosyncrasies) can be analysed via LS.[6]

[4] Constrained Lexical Sharing is (implicitly) adopted also in Lowe (2015d).

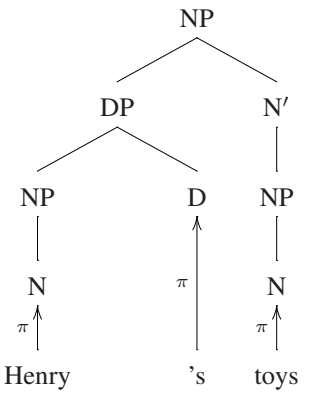
[5] It is also compatible with alternative approaches, such as that of Bögel (2012, 2015).

[6] It is also necessary to constrain LS in another respect, so that only sequences for which there is clear evidence for two c-structure nodes are analysed using LS. If unconstrained, potentially any morphologically complex word could be associated with multiple terminal c-structure nodes, which is clearly undesirable. See also fn. 23.

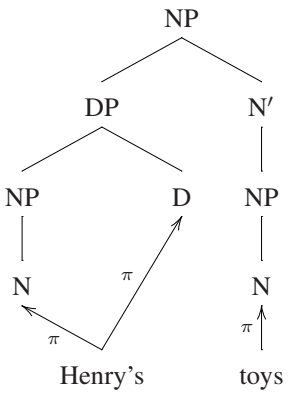
2.3 PDE possessive - formal analysis

When it comes to analysing a phenomenon like the PDE possessive, there are in principle three c-structural possibilities. As a full clitic, possessive *'s* would be a fully distinct lexical element from its host; it would constitute a separate element in the s-string, and map to its own terminal node in the c-structure, as in (6). As a full affix, possessive *'s* would not be distinct from its host: host and affix would constitute a single element in the s-string (and in the lexicon), and would map to a single terminal node in the c-structure, as in (8). The third possibility is a *lexically shared affix*: possessive *'s* would be an affix in the sense that the host-possessive sequence would constitute a single element in the s-string and lexicon, but this single element would map to two distinct terminal nodes in the c-structure, effectively granting the possessive marker a limited degree of independent status, at least in the c-structure. This last possibility is illustrated in (7).[7]

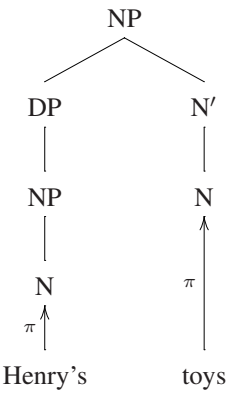
(6) Clitic:



(7) LS affix:



(8) Affix:

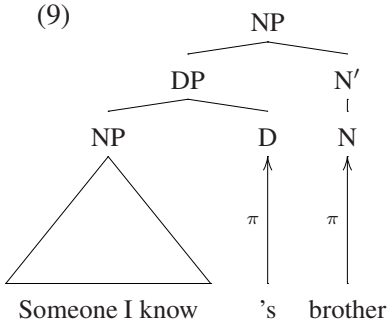


It is necessary to assume both a clitic and a lexically shared affix to account for the PDE data. Essentially, the 'regular' non-zero form of the possessive is a clitic: its distribution is so free, and its realization so regular, that it must be treated as an independent element in the lexicon, and therefore in the s-string and c-structure. Therefore (6) is the correct analysis of *Henry's toys*. The same is true, of course, when the possessive appears at the end of a postmodified possessor phrase, as in (9), for (3b). There is therefore no need to assume that all finite verbs, prepositions,

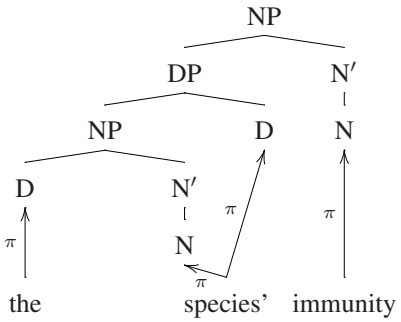
[7]The assumptions regarding noun phrase structure made in this paper are argued for in detail in Lowe (2015c). Essentially I do not follow the mainstream DP hypothesis approach to the English noun phrase, treating the maximal projection (ignoring quantifiers and numerals, at least) as NP, and the determiners as non-projecting D̂s appearing in Spec,NP. The possessive phrase, however, is analysed as a DP (the only DP in the language), with the possessive marker supplying the functional head of the phrase. On some level the DP label is arbitrary: it could equally be treated as a PossP, headed by Poss, or something similar. However, assuming D may have certain advantages: it permits a better integration with the assumptions of the traditional DP hypothesis, and as we will see it assists with explaining the diachronic development of the possessive marker. The non-standard assumptions made here are not crucial to the analysis, which could easily be reformulated under the standard DP hypothesis (as in Lowe 2013).

etc., have lexically specified 'possessive' forms. However, the 'zero' possessive can only be analysed affixally. Although a form like *species'* must be treated as a lexical unit, it nevertheless patterns with two-word sequences involving the clitic possessive *'s*. It is therefore best analysed via Constrained Lexical Sharing: a single element in the s-string (and lexicon) projects to two terminal nodes in the c-structure; see (10). This results in a uniform c-structure for all possessive phrases in PDE, despite the fact that in some the possessive D node is filled by an independent word, and in others it is not.

(9)

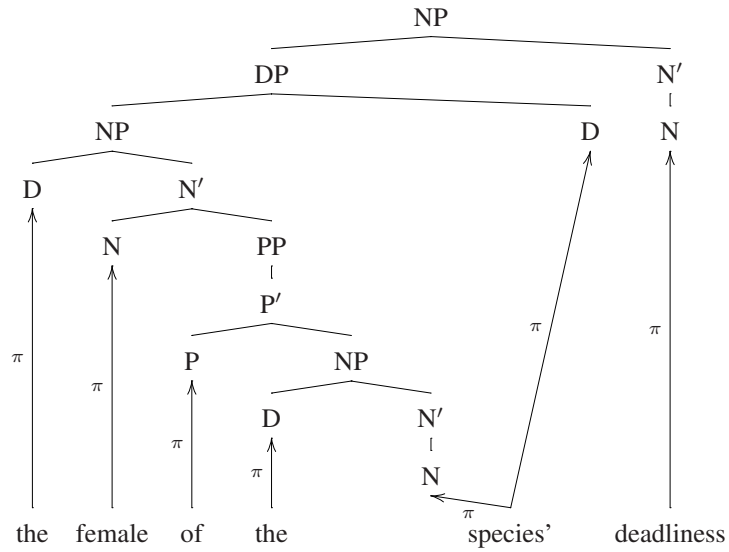


(10)



The lexical entry for the form *species'* (12) specifies that it fills two adjacent c-structure nodes, N and D. When a noun such as *species'* is used at the end of a postmodified possessor phrase, the analysis is the same (11). That is, the lexically specified possessive form of *species* (and similar nouns) is used when the 'next' node in linear terms is the D^0 node of the superordinate possessor DP, regardless of the function of the N component within the possessor phrase.

(11)



In practice, this means that the two terminal nodes of the c-structure with which a word like *species'* is associated will not necessarily project the same f-structure (they will project the same f-structure only if the N component is the head of the NP complement of the possessor DP, i.e. if the N is the head of the possessor phrase). The consequence of this is that the lexical entries for *species'* and parallel forms must be partitioned, such that all f-descriptions specified in the lexical entry are assigned to one or the other c-structure node. The f-description (POSS ↑) must be assigned to the D node, in order to constrain the possessor phrase to functioning as a possessor phrase at f-structure.[8] We must therefore assume a lexical entry of the form in (12) for *species'*, with all f-descriptions associated with either the N or the D. This is not the only possibility with LS, as we will see in the following section, but it is what we must assume for PDE. I will refer to this type of LS as *Partitioned Lexical Sharing*.[9]

(12)

species':	N	D
	(↑PRED) = 'species'	(POSS↑)

3 Diachrony

The origins and development of the English possessive have been discussed in detail by Allen (1997, 2003, 2008), and most of the data presented in this section is taken from her work. The development of the English possessive is of particular interest from a diachronic perspective as it appears to represent an example of the phenomenon of 'degrammaticalization', specifically an example of an affix developing into a clitic (cf. e.g. Norde 2012).[10]

Grammaticalization as a diachronic syntactic process has been observed to be largely unidirectional, and this unidirectionality has been claimed to be a defining, even exceptionless feature of it.[11] The possibility of the converse process, 'degrammaticalization' (Ramat 1992), is therefore controversial, and has been the subject of considerable interest.[12] Haspelmath (e.g 2004) argues forcefully that grammaticalization is unidirectional, and he shows that many supposed examples of degrammaticalization can be better analysed in different ways. Even so, Haspelmath accepts that this unidirectionality is merely a strong tendency rather than

[8]The POSS function is specified in the phrase-structure rule which introduces the possessive DP in the specifier position of NP.

[9]The simple reference in this lexical entry (and those provided below) to two category labels glosses over a more complex specification. The category labels in a lexical entry are abbreviations for functions such as '$\lambda(\pi(\bullet))$ = N'; in a lexical entry that specifies two category labels, there must be an additional specification constraining the order. The details of this are not important for the present purposes.

[10]The claim that the affix *-es* was reanalysed as an independent pronoun *his*, and subsequently underwent grammaticalization to a clitic, has been convincingly refuted by Allen.

[11]On grammaticalization in general see e.g. Heine (2003) and Hopper and Traugott (2003).

[12]See, for example, Campbell and Janda (2001) and other papers in *Language Sciences* 23, and also e.g. Geurts (2000), van der Auwera (2002), and Norde (2009, 2010).

an absolute rule. In particular, Haspelmath considers the English (and Mainland Scandinavian) change of genitive affix *-s* to possessive clitic to be one of the few genuine counterexamples to the unidirectionality of grammaticalization.[13]

In analysing the PDE possessive as a clitic, at least for the most part, I necessarily assume, with Haspelmath and others, a degrammaticalization.[14] However, this degrammaticalization is considerably more complex than a simple change of affix to clitic, as can be seen even from the PDE data, where both a clitic and an affix must be assumed.[15]

When it comes to tracing the degrammaticalization of the English possessive, we are faced with the usual difficulties that arise when studying 'dead' languages. The evidence at our disposal is an incomplete, written record of a dialectally diverse language which undoubtedly changed in different ways and to different extents in different dialect areas over the course of the period. That is, what a particular text tells us about a particular stage of the language as a whole can only be properly understood when set in the context of the text's composition; the problem is that some or all of this context may be lost to us. Nevertheless, at least in broad terms the evidence from Old English to PDE can be seen to reflect the development of a clitic possessive construction alongside the inherited affixal construction, via a LS construction.

3.1 Old English

What is now at least partly a clitic possessive marker clearly began life as an affix, appearing in Old English (OE: 7th–11th century A.D.) as a genitive case morpheme. OE was a relatively inflectional language compared with PDE (more similar to Modern High German), distinguishing up to five different cases in the declensional system. OE *-es*, the ancestor of PDE *'s*, was just one of a number of genitive case allomorphs suffixed to nouns and adjectives, distributed largely according to declensional class.

In addition, a number of noun and adjective stems showed not fully predictable idiosyncrasies in form when combined with their genitive suffix. For example, the noun *hnutu* 'nut' optionally shows umlaut of the stem in the genitive singular *hnyt-e*/*hnut-e*; stems ending in /h/ regularly drop that final segment outside the nom./acc.sg., and can show idiosyncrasies in how the case suffixes combine with the stem, e.g. *scōh* 'shoe, gen.sg. *scō-s*, gen.pl. *scō-na* (reflecting **scōh-ena*).

The genitive forms of nouns were lexically specified; adjectives uniformly adopted one of two declensional patterns depending on syntactic context. Pronouns

[13] For an analysis of Swedish possessive *-s* which rejects degrammaticalization, see Börjars (2003).

[14] Börjars et al. (2013) argue that it is not possible to treat the change in status of the possessive as a degrammaticalization, treating the variation in terms of a tension between head placement and edge placement, but not specifically discussing the clitic/affix question. As I argue above, and in more detail in Lowe (2015c), it is necessary to assume a clitic in at least some contexts in PDE.

[15] Note also that my interest in *grammaticalization* and *degrammaticalization* is restricted to its major consequences in terms of the grammatical categories and syntactic structure of the forms involved, and not more subtle grammatical changes, as discussed e.g. by Szmrecsanyi (2013).

also made case distinctions. All declinable words in a noun phrase were fully declined in the appropriate case, with agreement between modifers and nouns modified. So, in (13), both elements of the discontinuous noun phrase *manna gōdra* 'of good men' appear in the genitive singular:

(13) *man-na* *dǣda* *gōd-ra*
man-GEN.PL deeds good-GEN.PL
'the deeds of good men'

Altogether, genitive case marking in Old English, and the genitive morpheme *-es* in particular, display all the characteristics of inflectional affixation, and there is no reason to assume any other possibility at this stage of the language.

3.2 Early Middle English

By Early Middle English (EME: c. 1100–1400 A.D.) many of the inflectional features of the genitive were disappearing. The other cases had largely been lost, as had most variation in form of the old genitive: most nouns now formed their genitive/possessive in *-(e)s* (sg.), *-e* (pl.), but some still showed lexical irregularities, being somewhat similar, therefore, to the PDE (non-possessive) plural, where *-s* is predominant but by no means universal. At the same time, adjectival agreement began to drop out of use, and genitive/possessive marking began to be confined to the head noun. So, in (14) only the second noun in the phrase *Laferrd Cristess* is marked as genitive/possessive. Quantifiers retained agreement with nouns for longer, and retained inherited genitive forms for longer, even in texts where the marker was largely standardized as *-s* (15).

(14) *þe* *Laferrd* *Cristess* *bisne*
The Lord Christ.GEN example
'The Lord Christ's example.' (*Ormulum*, c. 1180 A.D.)

(15) *For* *ȝho* *iss* *all-re* *shafft-e* *cwen*
For she is all-GEN.PL creature-GEN.PL queen
'For she is queen of all creatures.' (*Ormulum*, c. 1180 A.D.)

Crucially, at this stage of Middle English (prior to 1400) we find two constructions which together suggest the beginnings of a change in the morphosyntactic status of the possessive marker. Firstly, when a possessor is a coordinated phrase, the possessive marking appears only once, on the rightmost head of the possessor phrase.

(16) *wif* *&* *weres* *gederunge*
wife and man.GEN union
'The union of man and wife.' (*Hali Meidenhad*, c. 1225 A.D.)

Secondly, a possessor phrase with postmodification would most commonly appear 'split', with the possession marked on the head of the possessor phrase, directly preceding the possessum, and the postmodifier appearing after the possessum. Allen (2013) attributes the beginnings of this construction to the late OE period, and argues that it originated via extraposition of the modifying phrase to the clausal right edge.

(17) *þe eorles douȝter of Gloucetre*
the earl.GEN daughter of Gloucester
'The Earl of Gloucester's daughter.' (*Polychronicon* VIII, c. 1380)

Common to both these constructions is a strong positional constraint, requiring that the possessive marker appearing on the head of the possessor phrase must immediately precede the possessum.[16] This constraint can be dealt with by assuming that the possessive marker is no longer fully affixal, in the sense that its position in the phrase is not simply dependent on the position of the word to which it attaches, but is also constrained by the context in which it appears. What this means in formal terms will be discussed in detail below.

3.3 Late Middle / Early Modern English

By the Late Middle English (LME) and Early Modern English (EModE) period (c. 1400–1600 A.D.), nearly all nouns use *-es* as the genitive/possessive marker, both singular and plural. Some nouns show morphophonologically 'irregular' (i.e. lexically specific) genitive forms, in considerably (and increasingly) smaller numbers than in EME, but in somewhat greater number and variety than in PDE (where the only morphophonologically irregular possessive is the 'zero' marked possessive, as discussed above).

In discussing PDE above, I took the existence of the 'phrasal possessive' as clear evidence for the clitic status of the possessive. It is therefore in the development of the phrasal possessive that the degrammaticalization of the possessive marker can be most clearly discerned (although, as discussed below, it is not so clear cut that we can definitely assume a clitic as soon as the phrasal possessive appears). The beginnings of the phrasal possessive may lie in the sorts of examples seen in (14) and (16), where the possessive appears only on one (the final) part of the head.[17]

Such examples show phrasal possession of a highly restricted kind; the possessive marking is still necessarily on the head (or one of the heads), even if one of the heads / part of the head is unmarked. The crucial change that occurred around the

[16] The other way of stating the constraint, namely that the possessive marker, which immediately precedes the possessum, must appear on the head, is descriptively accurate but diachronically unhelpful, since at this period it is appearance on the head of the possessor phrase, and not position directly before the possessum, which is inherited as an obligatory constraint.

[17] The development of the phrasal possessive is discussed in detail by Rosenbach (2004).

end of the 14th century involves the appearance of phrasal possessives proper, that is, phrasal possessives with postmodified possessors, where the possessive marking appears not on the head of the possessor but on the right edge of the postmodifying phrase. This is first found with relatively fixed, potentially lexicalized expressions, as in (18), but was increasingly freely formed with any such postmodified noun phrase.

(18) *The grete god of Loves name*
'The great God of Love's name.' (Chaucer, HF 1489, c. 1400 A.D.)

This is comparable to examples such as (3a); however, it is not in itself sufficient to demonstrate the existence of a clitic at this period with quite the same certainty as can be done for PDE, because such examples always (to my knowledge) involve postmodifiers that end with nouns. That is, while the possessive marking does not appear on the head of the possessor phrase, its distribution is still restricted to nouns. It could therefore quite reasonably be treated as an edge-affix without entailing an unacceptable degree of lexical duplication. On the other hand, the lack of examples parallel to (3b–d) is not in itself conclusive, since the phrasal possessive is itself relatively rare, and postmodified phrases most usually end in a noun anyway, so the gap might be accidental. Either way, the development seen here is clearly a step beyond the situation in EME, on the way towards the reanalysis of the possessive marker as a clitic.

During the same period, however, we also find apparently more affixal constructions. The type illustrated in (19) is most clearly affixal: the marking appears on the head of the possessor and within the possessor phrase, rather than, as is usually the case, at the end of the possessor phrase immediately before the possessum. This is the only construction type in Middle English in which the positional constraint requiring the possessive marker to directly precede the possessum is violated. The 'split' possessive also remains a possibility during this period, as in (20), which is equivalent to (17).

(19) *My lordes of Suffolk men*
'My Lord of Suffolk's men.' (Paston II, c. 1450 A.D.)

(20) *The Wyves Tale of Bathe*
'The Wife of Bath's Story.' (Chaucer, CT D, c. 1400 A.D.)

The strategies illustrated in (18) and (20) for expressing possession on a postmodified noun phrase exist side by side throughout the LME/EModE period; the strategy in (19), the most affixal, also coexists with these for some time, though it is less common, and dies out earlier. Throughout this period there is a gradual change in preference between the construction in (20) and that in (18). Allen (2013: 18) gives the following data for the change.

(21)

Period	Ex. (20) type (%)	Ex. (18) type (%)
1150–1250 (M1)	100	0
1250-1350 (M2)	100	0
1350-1419 (M3)	86	14
1420-1499 (M4)	46	54
1500-1569 (E1)	5	95
1570-1639 (E2)	9	91
1640-1710 (E3)	12	88

These two constructions therefore existed in competition alongside one another for a considerable time, with a gradual change in preference towards the less affixal, more clitic-like construction. The fully affixal construction (19) survives into EModE: the latest Allen (2013) cites is from the E1 period (22). But the same author also uses a phrasal possessive (23).

(22) *the bishopps of Rome perswations*
'The bishop of Rome's arguments.' (Wyatt, E1 period letter)

(23) *The Duke of Cleves man that is here*. (Wyatt, E1 period letter)

As Juvonen (2013) notes, at this period the possessive could no longer be analysed as purely inflectional, but also could not be considered unproblematically a clitic. Clearly we cannot talk about an absolute or sudden change from affix to clitic (as assumed, for example, by Carstairs 1987), but must recognize a gradual process of change over time. That is, there was a period in which a 'less affixal' phrasal possessive construction could be used with certain types of possessor phrase, while with other possessors the 'more affixal' possessive construction was possible. For example, the phrasal possessive was established as a possibility at a time when more clearly affixal 'genitive' forms were still found. Altogether we see a gradual development whereby the use of the possessive marker changes in such a way that by the EModE period it is at least close to becoming possible to reanalyse it as a clitic. We also see multiple constructions co-existing alongside one another, such that it would not be unreasonable to assume that at any one time, even for single speakers, there were multiple possible ways of analysing the possessive.

In the following section, I propose how this complex diachronic development may be modelled within the framework of Lowe (2015c), in such a way that each stage can be seen to follow on from the next one in a broad movement from unambiguous affix towards clitic.

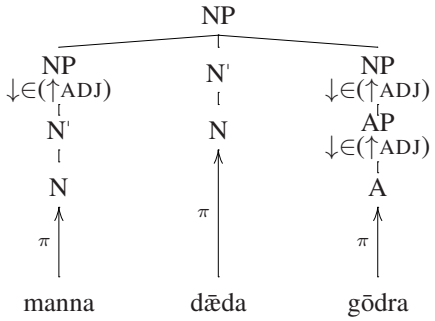
4 Historical development of the possessive

4.1 Old English

As discussed in §3.1, possession in OE is expressed by means of the genitive case, and there is no evidence for treating this as anything other than a fully affixal phe-

nomenon. As such, the genitive affixes, including *-es*, the ancestor of PDE *'s*, must be treated as fully integrated with their stems, both in the s-string and the c-structure (as with the affixal analysis of *Henry's* in (8)). This is illustrated in (24) for the sentence in (13).[18]

(24)



4.2 Early Middle English

As discussed above, in the EME period many of the features that made the affixal status of the OE possessive marker unambiguous were lost. The increasing restriction of possessive marking to the final element of the possessor phrase, as long as this element was a head, does not necessarily mean that any change need be assumed in the status of the possessive marker itself, merely in the rules of possessive marking. However, the strong constraint on the possessive marker, that it should appear directly before the possessum, does mark the first steps towards the separation of the possessive marker from its host, since its position could no longer be stated purely by reference to the position of its host.

It is notable that by the EME period, the definite article was developing or had already developed, and possessive phrases had come to mark the definiteness of the possessum. The latter development can be dated to the end of the OE period, while the former is more difficult to pinpoint, but was most likely underway by the end of the OE period.[19] I therefore propose that the positional constraint on the possessor head be analysed by assuming the reanalysis of the increasingly frequent determiner node D as part of the possessor phrase, once the presence of a possessor phrase served to supply inherently definite reference to the possessum. That is, once possessor phrases began to necessarily express the definiteness of the possessum, they were fulfilling the function of an explicit determiner, and so came to be associated with the node within the noun phrase in which a determiner would appear. This reanalysis led to the reanalysis of possessive phrases as DPs. There was no explicit determiner to fill the head D, but the corresponding functional contribution was made by the head noun, marked with the possessive affix. It was then possible to reanalyse a fully affixal form like *eorles* in (17) as relating

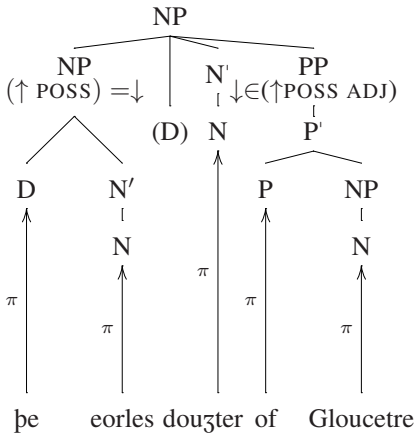
[18] Since there is no specialized possessive phrasal structure at this period I represent the possessive adjunct as an f-structure ADJ, contrasting with POSS in (25) below, but the point is moot.

[19] See e.g. Wood (2003, 2007a,b).

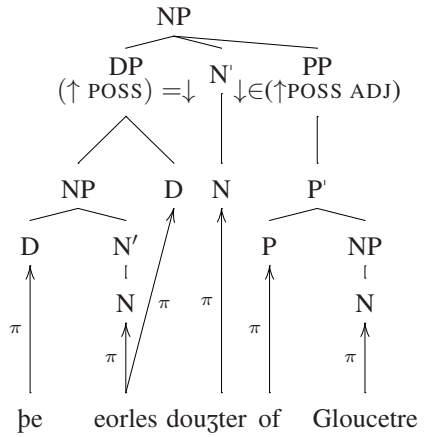
to two nodes in the c-structure: the head N of the possessor, and the D head of the possessor phrase. The motivation for the reanalysis was the structure of non-possessed noun phrases, which increasingly frequently took a definite marker in specifier position. By reanalysing the possessor phrase as a DP in Spec,NP, a more uniform NP structure covering both possessed and non-possessed phrases, with a definiteness-marking element in Spec,NP, was obtained. This reanalysis very simply accounts for the positional constraint: given the position of the D node directly preceding the core noun phrase, it was not possible for anything to intervene between a possessive-marked noun and its possessum, once the possessive marker became associated with D.[20]

That is, I propose that an original structure such as in (25), parallel to (24) except for the presence of an optional D node, was reanalysed as in (26), effectively by incorporation of the D into the head of the possessor phrase.

(25)



(26)



The strict positional constraint on the possessive marker cannot be stated of the head noun itself, since there is no requirement for the head of a noun phrase to appear at the right edge of its phrase; it is only when the noun appears with the possessive marker that the constraint applies, and in addition the form must also directly precede the possessum. This is very easily accounted for by assuming an association between possessive-marked noun and a functional head in the c-structure which appears directly after the possessor phrase and directly before the possessum.[21]

[20] An alternative suggestion, from Oleg Belyaev (p.c.), is that loss of case marking, except for *-es*, could have left *-es* looking like a separate element, and therefore liable to be reanalysed as the head of a functional phrase. At the least, this may have provided further support for the reanalysis proposed here.

[21] An alternative way of dealing with this positional requirement might be to make use of non-projecting categories (Toivonen 2003), along similar lines to Arnold and Sadler's (2013) proposals regarding PDE prenominal modification. Superficially, at least, the constraint against postmodification with possessives looks very similar to the constraint against postmodification with PDE prenominal modifiers. The details are slightly different, however; crucially, coordination of possessor-marked

Although *eorles* in (26) exhibits LS and appears entirely parallel to *species'* in (10) and (11), it is subtly different, in that the noun component must supply the head, or one of the heads, of the possessor phrase. This can be enforced via an f-description associated with the N component of the lexical entry, requiring that the f-structure projected from the noun serve as the value, or one of a set of values, of the attribute POSS in a superordinate f-structure. But this in turn means that there is no need to associate any f-descriptions with the D component of the lexical entry: the only f-description that needs to be associated with the D node is the constraint to ensure that the phrase functions as a POSS, so if that can be associated with the N, the D need have no f-descriptions specifically associated with it. In this way, the kind of LS required for *eorles* here is more unified than that required for PDE *species'* above (12): the f-descriptions contained in the lexical entry are not divided, but all associated with the same c-structure node. Such a lexical entry differs from that of an 'ordinary' word (i.e. one that does not display LS of any sort) only in that there is a second c-structure node with which the form is associated, although no functional information is associated with that node. This kind of LS can be labelled *Unified Lexical Sharing*, and distinguished from the *Partitioned Lexical Sharing* seen with PDE *species'* above.[22] I therefore assume a lexical entry of the following sort for EME *eorles* as seen in *þe eorles douȝter of Gloucetre*:

(27)

eorles:	N	D
	(↑PRED) = 'earl'	
	(POSS(∈) ↑)	
	((POSS ↑) DEF) = +	

We can interpret the difference between Unified Lexical Sharing and Partitioned Lexical Sharing as different degrees of degrammaticalization between affix and clitic. The form *species'* is in a sense closer to being two distinct forms, and correspondingly further from being an unambiguously unitary sequence, than *eorles*, since its lexical entry is effectively partitioned in two, while *eorles* presents a unified lexical entry which merely maps to two c-structure nodes. In this way the (phonologically null) possessive marker in *species'* is closer to a clitic than the possessive marker on *eorles*. The form *species'*, just like a sequence of two separate words, involves two c-structure nodes each with their own set of f-descriptions, with the division unambiguously correlating with the division between noun stem and possessive affix; the only difference between this and two separate words is that *species'* is a single unit in the s-string. On the other hand, the only difference between *eorles* and an 'ordinary' single word is that it projects to two nodes in the c-structure.[23]

nouns would be predicted to be obligatory, which it is not. In addition, a non-projecting analysis does not help with accounting for the diachronic development.

[22] In Lowe (2015c) I labelled these 'Head-Marked Lexical Sharing' and 'Phrase-Marked Lexical Sharing' respectively, but the current labels are preferable.

[23] Since all nodes are in principle optional, it could be argued that LS ought necessarily to involve

4.3 LME/EModE

As discussed above, the crucial change in this period was the development of the phrasal possessive with postmodification, as in (18). This requires at least a Partitioned Lexical Sharing analysis: the f-description constraining the phrase to function as a possessor must be associated with the D node, since the N node may be embedded within the possessor and may not itself be the head of the possessor. The alternative is a full clitic analysis, at least for forms like *loves* where there is no morphophonological idiosyncrasy in the form of the possessive marker. However, as stated above, it is not absolutely necessary to assume that we are dealing with a clitic at this stage, since there is no positive evidence of the attachment of the possessive marker to words other than nouns, and it is this that I take as crucial evidence for a clitic in PDE in Lowe (2015c).

At the same time, as noted above, a construction which can only be analysed as fully affixal, i.e. the type in (19) and (22), continues to be possible. We also see a gradual change in preference between the Unified Lexical Sharing analysis and the Partitioned Lexical Sharing analysis, from the former to the latter.

4.4 Overview of the diachronic development

For OE there is no need to assume anything other than a simple affix. In the EME period, evidence for structures involving Unified Lexical Sharing emerge, but constructions that cannot be so analysed and rather require a simple affixal analysis are also found. In the LME/EModE period evidence for Partitioned Lexical Sharing emerges, and over a period of a few hundred years becomes increasingly popular in comparison with the structures requiring a Unified Lexical Sharing analysis. In this period the simple affixal construction also exists, though it eventually drops out of use. It is possible that some structures requiring a full clitic analysis also develop in the LME/EModE period, although this is uncertain. In PDE the main possibilities are a full clitic and Partitioned Lexical Sharing, although the marginal continued existence of the 'split' possessive in PDE (Denison et al. 2010, Börjars et al. 2013) means that Unified Lexical Sharing may also be available for some speakers (Lowe 2015c: 29). This diachronic development is summarized in (28).

---

two terminal nodes that, at least potentially, map to *different* f-structures, such that there ought necessarily to be functional annotations associated with each node of a lexically shared word. That is, if a single word contributes information to only one f-structure, there is no functional motivation for it to be associated with more than one node in the c-structure. However, the claim here is that there may also be a *c-structural* motivation for a single word to map to two terminal nodes; in this case, the motivation is the strict positional constraint on possessive-marked nouns. So, Partitioned Lexical Sharing is an appropriate analysis for words for which there is functional, and perhaps also structural, evidence for association with two terminal nodes, whereas Unified Lexical Sharing is appropriate for words for which there is only structural evidence for such an association.

(28)

	Affix	LS affix (Unif.)	LS affix (Part.)	Clitic
OE	✓			
EME	✓	✓		
LME/EModE	(✓)	✓	✓	(✓)
PDE		(✓)	✓	✓

Overall, the development can be seen to involve degrees of degrammaticalization between affix and clitic; the development is complex and gradual, involving synchronic variation between more or less affixal/clitic-like structures at all points during the period of change.[24]

5 Conclusion

The English possessive is a more complex phenomenon than a simple clitic/affix distinction can capture, both synchronically and diachronically. In this paper, I have shown that LFG's Lexical Sharing makes it possible to represent the degrees of degrammaticalization shown by the English possessive at different periods. The development is from affix towards clitic, but through intermediate steps of LS constructions. This analysis crucially enables us to maintain the absolute distinction between word and morpheme required by a lexicalist theory.

References

Allen, Cynthia L. 1997. The Origins of the 'Group Genitive' in English. *Transactions of the Philological Society* 95(1), 111–131.

Allen, Cynthia L. 2003. Deflexion and the development of the genitive in English. *English Language and Linguistics* 7(1), 1–28.

Allen, Cynthia L. 2008. *Genitives in Early English: Typology and Evidence*. Oxford: Oxford University Press.

Allen, Cynthia L. 2013. Dealing with postmodified possessors in early English: Split and group genitives. In Kersti Börjars, David Denison and Alan Scott (eds.), *Morphosyntactic Categories and the Expression of Possession*, pages 1–34, Amsterdam: Benjamins.

Anderson, Stephen R. 2008. The English "Group Genitive" is a Special Clitic. *English Linguistics* 25, 1–20.

Arnold, Doug and Sadler, Louisa. 2013. Displaced Dependent Constructions. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG13 Conference*, pages 48–68, Stanford, CA: CSLI Publications.

[24]It is a degrammaticalization in terms of grammatical independence, in terms of the gradient between word and morpheme, but it does not necessarily involve change in terms of the cline between lexical and functional elements, since at all periods the contribution of the possessive marker is functional.

Asudeh, Ash. 2012. *The Logic of Pronominal Resumption*. Oxford: Oxford University Press.

Bögel, Tina. 2012. The P-Diagram – A Syllable-Based Approach to P-Structure. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG12 Conference*, pages 99–117, Stanford, CA: CSLI Publications.

Bögel, Tina. 2015. *The Syntax-Prosody Interface in Lexical Functional Grammar*. Ph. D.thesis, Universität Konstanz.

Börjars, Kersti. 2003. Morphological status and (de)grammaticalization: the Swedish possessive. *Nordic Journal of Linguistics* 26(2), 133–163.

Börjars, Kersti. 2013. Modelling linguistic change: grammaticalization, gradualness and persistence. Keynote lecture at the 21st International Conference on Historical Linguistics, University of Oslo, 6th August 2013.

Börjars, Kersti, Denison, David, Krajewski, Grzegorz and Scott, Alan. 2013. Expression of Possession in English. In Kersti Börjars, David Denison and Alan Scott (eds.), *Morphosyntactic Categories and the Expression of Possession*, pages 123–148, Amsterdam: Benjamins.

Campbell, Lyle and Janda, Richard. 2001. Introduction: Conceptions of Grammaticalization and their Problems. *Language Sciences* 23(2–3), 93–112.

Carstairs, Andrew. 1987. Diachronic Evidence and the Affix-Clitic Distinction. In Anna Giacalone Ramat, Onofrio Carruba and Giuliano Bernini (eds.), *Papers from the 7th International Conference on Historical Linguistics*, pages 151–162, Amsterdam: Benjamins.

Dalrymple, Mary and Mycock, Louise. 2011. The Prosody-Semantics Interface. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG11 Conference*, pages 173–193, Stanford, CA: CSLI Publications.

Denison, David, Scott, Alan K. and Börjars, Kersti. 2010. The real distribution of the English "group genitive". *Studies in Language* 34(3), 532–564.

Geurts, Bart. 2000. Explaining grammaticalization (the standard way). *Linguistics* 38(4), 781–788.

Haspelmath, Martin. 2004. On directionality in language change with particular reference to grammaticalization. In Olga Fischer, Muriel Norde and Harry Perridon (eds.), *Up and Down the Cline: The Nature of Grammaticalization*, pages 17–44, Amsterdam: Benjamins.

Heine, Bernd. 2003. Grammaticalization. In Brian D. Joseph and Richard D. Janda (eds.), *The Handbook of Historical Linguistics*, Oxford: Blackwell.

Hopper, Paul J. and Traugott, Elizabeth Closs. 2003. *Grammaticalization*. Cambridge: Cambridge University Press, second edition.

Juvonen, Teo. 2013. Variation in the form and function of the possessive morpheme in Late Middle and Early Modern English. In Kersti Börjars, David Denison and Alan Scott (eds.), *Morphosyntactic Categories and the Expression of Possession*, pages 35–58, Amsterdam: Benjamins.

Kaplan, Ronald M. 1989. The Formal Architecture of Lexical-Functional Grammar. In Chu-Ren Huang and Keh-Jiann Chen (eds.), *ROCLING II: Proceedings of the Computational Linguistics Conference*, pages 3–18, Tapei: The Association for Computational Linguistics and Chinese Language Processing (ACLCLP).

Lowe, John J. 2013. English Possessive *'s*: Clitic *and* Affix. *LSA Annual Meeting Extended Abstracts* 2013, `http://elanguage.net/journals/lsameeting/article/view/3524`.

Lowe, John J. 2015a. *Participles in Rigvedic Sanskrit: The Syntax and Semantics of Adjectival Verb Forms*. Oxford: Oxford University Press.

Lowe, John J. 2015b. Clitics: Separating Syntax and Prosody. *Journal of Linguistics* FirstView article available on CJO, DOI: 10.1017/S002222671500002X. 45pp. Currently scheduled to appear in vol. 52(2), July 2016.

Lowe, John J. 2015c. English Possessive *'s*: Clitic *and* Affix. *Natural Language and Linguistic Theory* To appear.

Lowe, John J. 2015d. The Syntax of Sanskrit Compounds. *Language* 91(3), e71–e115.

Mycock, Louise and Lowe, John J. 2013. The Prosodic Marking of Discourse Functions. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG13 Conference*, pages 440–460, Stanford: CSLI Publications.

Norde, Muriel. 2009. *Degrammaticalization*. Oxford: Oxford University Press.

Norde, Muriel. 2010. Degrammaticalization: Three Common Controversies. In Katerina Stathi, Elke Gehweiler and Ekkehard König (eds.), *Grammaticalization: Current Views and Issues*, pages 123–151, Amsterdam: Benjamins.

Norde, Muriel. 2012. Lehmann's parameters revisited. In Kristin Davidse, Bert Cornillie, Hubert Cuyckens and Torsten Leuschner (eds.), *Grammaticalization and Language Change: New Reflections*, pages 73–109, Amsterdam: Benjamins.

Payne, John. 2009. The English genitive and double case. *Transactions of the Philological Society* 107(3), 322–357.

Quirk, Randolf, Greenbaum, Sidney, Leech, Geoffrey and Svartvik, Jan. 1985. *A Comprehensive Grammar of the English Language*. Harlow: Longman.

Ramat, Paolo. 1992. Thoughts on degrammaticalization. *Linguistics* 30, 549–560.

Rosenbach, Anette. 2004. The English *s*-genitive: A case of degrammaticalization? In Olga Fischer, Muriel Norde and Harry Perridon (eds.), *Up and Down the Cline: The Nature of Grammaticalization*, pages 73–96, Amsterdam: Benjamins.

Szmrecsanyi, Benedikt. 2013. The Great Regression: Genitive variability in Late Modern English news texts. In Kersti Börjars, David Denison and Alan Scott (eds.), *Morphosyntactic Categories and the Expression of Possession*, pages 59–88, Amsterdam: Benjamins.

Toivonen, Ida. 2003. *Non-Projecting Words: A Case Study of Swedish Verbal Particles*. Dordrecht: Kluwer.

van der Auwera, Johan. 2002. More thoughts on degrammaticalization. In Ilse Wischer and Gabriele Diewald (eds.), *New Reflections on Grammaticalization*, pages 19–29, Amsterdam: Benjamins.

Vincent, Nigel. 1999. The evolution of c-structure: prepositions and PPs from Indo-European to Romance. *Linguistics* 37(6), 1111–1153.

Vincent, Nigel. 2001a. Competition and correspondence in syntactic change: Null arguments in Latin and Romance. In S. Pintzuk, G. Tsoulas and A. Warner (eds.), *Diachronic Syntax: Models and Mechanisms*, Oxford: Oxford University Press.

Vincent, Nigel. 2001b. LFG as a Model of Syntactic Change. In Miriam Butt and Tracy Holloway King (eds.), *Time over Matter: Diachronic Perspectives on Morphosyntax*, pages 1–42, Stanford, CA: CSLI Publications.

Wescoat, Michael Thomas. 2002. On Lexical Sharing. Ph.D. Thesis, Stanford University.

Wescoat, Michael Thomas. 2005. English Nonsyllabic Auxiliary Contractions: An Analysis in LFG with Lexical Sharing. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG05 Conference*, pages 468–486, Stanford, CA: CSLI Publications.

Wood, Johanna L. 2003. *Definiteness and number: Determiner phrase and number phrase in the history of English*. Ph. D.thesis, Arizona State University.

Wood, Johanna L. 2007a. Demonstratives and possessives: From Old English to present-day English. In Werner Abraham, Elisabeth Leiss and Elisabeth Stark (eds.), *Nominal determination: Typology, context constraints, and historical emergence*, pages 337–359, Amsterdam: Benjamins.

Wood, Johanna L. 2007b. Is there a DP in Old English? In Joe Salmons and Shannon Dubeenion-Smith (eds.), *Historical Linguistics 2005: Selected Papers from the 17th International Conference on Historical Linguistics, Madison, Wisconsin, 31 July – 5 August 2005*, pages 167–187, Amsterdam: Benjamins.

Zwicky, Arnold M. 1987. Suppressing the *Z*s. *Journal of Linguistics* 23, 133–148.

CLITIC POSITIONING IN OSSETIC

John J. Lowe
University of Oxford

and

Oleg Belyaev
Lomonosov Moscow State University

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King (Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

We present data regarding the positioning of clitics in the Iron dialect of Ossetic, which is in some respects similar to data for 'second-position' clitics in more familiar languages, but in other respects is considerably more complicated. We propose a formal analysis which makes use of Optimality Theoretic LFG with stochastic re-ranking of constraints, which is able to capture the full range of the data, while avoiding having to assume the controversial phenomenon of clitic 'movement'.

1 Introduction

The syntactic and prosodic properties of so-called 'second-position' clitics, and how these properties interact to determine their surface positioning, has been the subject of considerable work. Essentially, the question is to what extent the surface position of clitics in their clause can be accounted for by ordinary syntactic processes, and to what extent they must also be constrained and influenced by other factors, such as prosody. Accounts of clitic positioning in LFG include Newman (1996), Austin and Bresnan (1996), Nordlinger (1998a,b), O'Connor (2002a,b), Wescoat (2009), Bögel (2010, 2014, 2015), Bögel et al. (2010), and Lowe (2011, 2015a). In this paper we discuss the complex set of constraints on 'second-position' clitics in the Iron dialect of Ossetic.[1] Our position in this paper is that it is methodologically desirable to aim for a primarily syntactic account of a syntactic phenomenon — word order patterns — while acknowledging the influence of prosodic factors.[2]

The Iron Ossetic data regarding 'second-position' clitics have never been fully

[†] The authors are very grateful to the audiences at SE-LFG 17 and LFG15 for their attention, and in particular to Mary Dalrymple, John Payne, and Andy Spencer for valuable insights, and also the reviewers for their time and comments. This work was undertaken while the first author was in receipt of an Early Career Research Fellowship funded by the Leverhulme Trust. Oleg Belyaev is grateful to Madina Darchieva for grammaticality judgements on the Ossetic data. Belyaev's work has been supported by the Russian Foundation for the Humanities, project no. 13-04-00342.

[1] Ossetic is an Iranian language spoken by approximately 700,000 people, the majority of whom live in the Republic of North Ossetia–Alania, Russian Federation. Ossetic consists of two closely-related dialects, Iron and Digor, both of which have standard varieties, but Iron is by far the dominant dialect, and the basis of the standard Ossetic language. This study is based on the standard variety of the Iron dialect. The transcription used for Ossetic examples follows Dzaxova (2009), with the following exceptions: /ə/ is used instead of Dzaxova's /ɔ/; single letters are used instead of digraphs for affricates (/c/ for /t^{s}/, /č/ for /t^{š}/, /ž/ for /d^{ž}/). Geminate consonants are written as two letters if they belong to two different morphemes (e.g. *kod-t-on* (do-TR-PST.1SG), phonologically /kot:on/), and with the length mark otherwise (*dɜtː-ən* (give-INF) 'to give'). Most examples of Ossetic sentences given in this paper are taken from the Ossetic National Corpus (http://corpus.ossetic-studies.org/en, c. 5 million wordforms); other examples are taken from published work, or from the fieldwork of the second author.

[2] In his analysis of Digor Ossetic clitics, Erschler (2010) shows that apparently prosodic positioning in that dialect is not, in fact, purely prosodic.

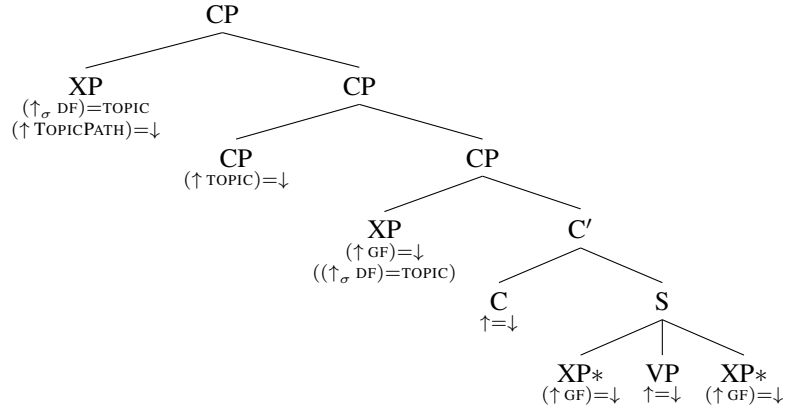
described, nor subject to formal analysis.[3] Much of the data regarding clitic positioning in Ossetic are very similar to data found in other languages, and can be analysed along similar lines. However, some data are, to our knowledge, unparalleled in the wider literature.

2 Ossetic clause structure

There is broad agreement on the hierarchical structure of the Ossetic clause.[4] For this paper, the finer details are not important; what matters is the generally agreed-upon right-branching hierarchical structure, with a number of positions in the left periphery associated with particular information structure roles. Ossetic word order is to a large extent 'discourse-configurational'. The unmarked order of constituents is SOV, but in principle all possible orderings of S, O and V can be grammatical. The constraints on the choice of orderings are essentially information structural: broadly speaking, topics appear near the left-edge of the clause, before the complementizer position, while foci appear to the right of the complementizer position, and to the left of the verb complex. While the order of major constituents is relatively free, there are some parts of the Ossetic clause which are relatively fixed. These include the verbal complex, in which a number of relatively grammaticalized particles must appear, along with the verb (Erschler 2012).

We assume a phrase-structure very similar to that assumed by Belyaev and Haug (2014), but differing in a few details. The structure we assume is shown in (1) and (2); these structures represent the range of possibilities, but all (terminal) nodes are optional, that is not all will be filled in any one clause.

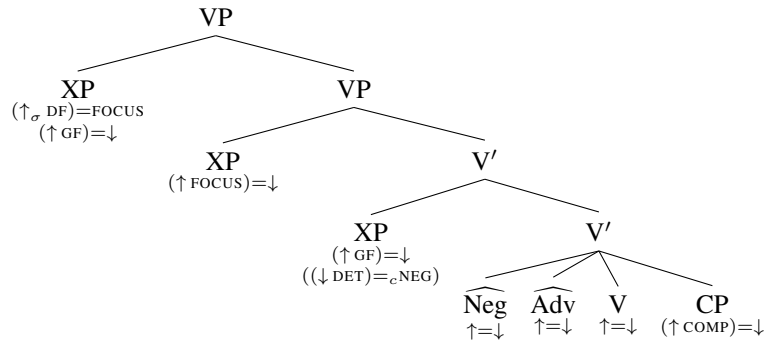
(1)



[3]Erschler (2010) provides a preliminary analysis of clitics in Digor Ossetic, but certain facts in that dialect are rather different from those in Iron.

[4]Compare Lyutikova and Tatevosov (2009) in the transformationalist tradition with e.g. Belyaev and Haug (2014) within LFG.

(2)



We treat the core of the Ossetic clause as a CP. Various elements may appear to the right or left of the core clause, however. Left-dislocated and right-dislocated topic phrases are adjoined to CP.[5] Relative clauses which form part of a correlative-relative construction are also adjoined to CP (the slot marked (↑ TOPIC) =↓).[6] The first element of the core clause is an optional XP appearing in Spec,CP. This phrase can fill any grammatical function in the main clause, but may not head a long-distance dependency. We assume here that XP may include DP/NP, AdjP, AdvP or PP, though the inventory of phrasal categories in Ossetic is debatable. One type of phrase may not appear in any of the XP slots in the trees: the VP is constrained to appear only in the VP slot (or, in the case of some non-finite verb phrases, embedded under a noun phrase). Following the first core position is a node, C, that can be filled by certain subordinating conjunctions (non-preverbal subordinating conjunctions: *kɜd* 'if', *səma* 'as if', *salənmɜ* 'while', *sɜmɜj* 'in order to', *jugɜr* 'if'). We then allow for the possibility of one or more phrases of any type to precede or follow the VP. Phrases that function as foci in information structure terms are adjoined to VP: there may in principle be any number of them, and they immediately precede the verbal complex, except that a few phrase types can intervene between IS foci and the verb complex. Of these, interrogative phrases, correlative phrases, and preverbal subordinators appear in Spec,VP, and negative phrases are adjoined to V′. The verbal complex consists of the V′, dominating the V, non-projecting negative particles and certain adverbs and adverbial particles which do not contribute PRED values, and embedded complement clauses.

3 Clitic positioning - the data

In Iron Ossetic all 'second-position' clitics in a clause must appear together in a single 'clitic cluster'.[7] There is a fixed order of elements within the clitic cluster,

[5] Right-dislocation is not shown in (1). On the annotations specifying information structure status, e.g. (↑σ DF) = TOPIC, see Dalrymple and Nikolaeva (2011). TOPICPATH resolves to (↑ COMP* GF), i.e. an argument or adjunct of the clause or any of its complements.

[6] We follow Belyaev and Haug (2014) in using f-structure TOPIC and FOCUS features to govern the correlative-relative structure, though the details are not important for this paper.

[7] Erschler (2010) shows that this is not the case in Digor, where clitics are in general freer in terms of their positioning.

but the details of this are unimportant for the present purposes. We treat this by means of the CCL node introduced by Bögel et al. (2010), which we treat as an exocentric category dominating any number of non-projecting CL nodes. Wherever below we refer to the positioning of a clitic, or clitics, we really refer to the positioning of a clitic cluster, which will include one or more clitics in any given clause.

(3) $\text{CCL} \rightarrow \underset{\uparrow=\downarrow}{\text{CL}^{+}}$

As a general rule, the 'second-position' occupied by 'second-position' clitics in the Ossetic clause can be unproblematically analysed in purely syntactic terms: clitics generally follow the first clause-level XP, regardless of its length.

(4) *=mɜm žawər =mɜm ɜrba-səd-i *=mɜm
me.ALL Zaur me.ALL PV-go-PST.INTR.3SG me.ALL
'Zaur came to me.'

(5) [žawər-ə *=dɜm rɜšuʁd *=dɜm čənz] =dɜm
Zaur-GEN thee.ALL beautiful thee.ALL bride thee.ALL
ba-zərd-t-a *=dɜm
PV-speak-TR-PST.3SG thee.ALL
'The beautiful bride of Zaur called for you.'

This is the first clause-level XP in the core CP. When a phrase is adjoined to the left of the core CP, clitics appear in *third* position:

(6) [ɜnuš-mɜ] [χɜrž-gɜn-ɜʒ-ə nom] =əl sɜmɜj š-bad-a
century-ALL good-do-PTCP-GEN name he.SUPER PURP PV-sit-SBJV.3SG
'So that he is forever called a good person.' (ONC)

(7) žon-əš, meret, [ɜž iron dɜrɜš-tɜ kɜj dar-ən],
know-PRS.2SG Meret I Ossetian clothing-PL COMP wear-PRS.1SG
[fəldɜr] =mɜ wəj təχχɜj nɜ waržə
more me.GEN that[GEN] for NEG love-PRS.3SG
'You know, Meret, he doesn't like me more because I'm wearing Ossetian clothes.' (ONC: *Max dug* 5, 1996)

In (6), the clitic follows the second full noun phrase; the first phrase is a left-dislocated constituent. In (7), the first phrase is a relative CP, while the initial XP of the main clause is the adverb *fəldɜr*, which is followed by the correlate *wəj*. The clitic follows the adverb, and thus it is technically in the third position; however, the correlative clause may be considered to be outside the core CP. Within the core CP, then, second syntactic position is the rule.

The situation is slightly more complicated if no XP appears to the left of the verbal complex.[8] If the main verb itself is the first word in the clause, the clitic cluster directly follows this:

[8]We omit here discussion of clitic positioning relative to complex predicates in Ossetic, and relative to clause-initial non-finite verb phrases.

(8) *nə-ffəšt-a* *=jɜ* *gardantə* *miχal*
PV-write-PST.3SG it.GEN G. M.
'It was Mikhal Gardanov who wrote it down.'

When the verb or the verbal complex falls into the scope of clitic cluster positioning, then, we see that for the most part the same principles apply that we have seen before: the clitic cluster appears in syntactic 'second position', which may be following the first full phrase or following the first word, where this is a lexical or functional head of the clause.[9] However, there is one context which cannot so easily be analysed, and which strongly suggests the influence of prosody on clitic positioning in Ossetic.

If the first elements in a clause are proclitics directly preceding the verb, the clitic cluster follows either the first proclitic, or the last proclitic immediately preceding the verb (9).[10]

(9) *sə* *(=šɜm)* *nɜ* *(=šɜm)* *wəd-i* **=šɜm,* *aχɜm* *nɜ-j*
what they.ALL NEG they.ALL be-PST.INTR.3SG they.ALL such NEG-is
'There is nothing that they did not have (lit. that to them there was not).' (spoken text)

In such examples, the first proclitic is always a proclitic question word. As such, this proclitic can be analysed as heading (and constituting) a full XP in the phrase-structure. Thus when the clitic appears following such a proclitic, we can analyse it in exactly the same way as all the examples we have seen above: the clitic follows the first full phrase of the core clause, which in this case merely happens to consist of a single proclitic word.[11] On the other hand, the possibility of the clitic appearing following more than one proclitic violates the generalizations we have made so far. As stated, in such a case the clitic appears directly before the verb, regardless of how many proclitics precede. In (9), the second proclitic is a negative particle which, following the phrase structure proposed in (2), is a sister to V. Under no analysis is it possible to treat the two proclitics preceding the verb as somehow forming a single constituent for the clitic cluster to follow.[12]

However we choose to analyse this alternation, it seems we must admit prosodic influence on the positioning of the clitic cluster, since the problematic position of

[9] In this, Ossetic clitics show similar distributional properties to clitics in some Slavic languages (e.g. Franks and King 2000).

[10] Ossetic proclitics are prosodically deficient insofar as they do not have lexical stress and require a host to their right on which to 'lean'. They are, however, positioned in the clause according to entirely 'ordinary' syntactic rules/processes, and as such their positioning can be accounted for entirely syntactically. We can therefore distinguish between elements that are *prosodically* clitic but not syntactically so (the 'simple clitics' of Zwicky 1977), and elements that are *syntactically* clitic (and also, though not necessarily, prosodically clitic), in Ossetic enclitics. In this paper we use the term *clitic* to refer exclusively to elements that are syntactically clitic, regardless of prosodic features. This means that our term *clitic* in relation to Ossetic refers only to elements that are enclitic.

[11] There is nothing typologically unusual about proclitics being able to host enclitics; this is also clearly attested in Ancient Greek, for example.

[12] Note also that it is not possible for anything to appear between proclitic particles and the V inside the verbal complex, apart from the clitic cluster in contexts such as that in (9).

the clitic cluster is possible only when the first elements in the clause are all proclitic.

Another phenomenon which provides further evidence for prosodic influence on clitic positioning is so-called 'clitic climbing'. Superficially, Ossetic displays two different kinds of clitic climbing, which we call 'strong clitic climbing' and 'weak clitic climbing'.

Strong clitic climbing (SCC) corresponds to clitic climbing in e.g. French, where clitics can appear in a higher clause, in c-structure terms, from the clause in which they must be understood. As in Romance languages, this can be taken as evidence for a complex predicate, i.e. for the monoclausal status at f-structure of a construction involving two separate verb forms.[13] In Ossetic, clitics that 'climb' to a higher clause are subject to exactly the same positioning constraints as clitics in simple clauses, and as such SCC does not contribute any new data to the analysis of clitic positioning itself; we therefore do not consider it further here.

Weak clitic climbing (WCC) is another matter. In general, WCC is a third option alongside SCC and in-situ positioning for clitics in certain embedded structures. In WCC, the clitic appears at the left-edge of the embedded clause. It 'climbs' insofar as it appears in a position that is not possible for a non-embedded clause, and insofar as it is prosodically cliticized onto the last word of the matrix clause, but in purely phrase structural terms there is no need to assume that the clitic has 'climbed' out of its clause.[14]

(10) *mɜn* *fɜnd-ə* *[(=jɜm)* *alə* *bon* *(=ɜm)* *kɜš-ən]*
me.GEN want-PRS.3SG it.ALL every day it.ALL look-INF
'I want to watch it every day.'

(11) *a-sə* *tɜkːɜ* *šaχat-əl* *=mɜ* *qɜw-ə* *[(=jɜ)*
this-ATTR exactly hour-SUPER me.GEN need-PRS.3SG it.GEN
ba-žon-ən *(=ɜj)]*
PV-know-INF it.GEN
'It is exactly now that I need to know it.' (ONC: *Max dug* 9, 2007)

What is relevant here is that WCC shows us clitics appearing in a position within an embedded clause that they could not occupy in a non-embedded clause. Contrast the variable positioning of the clitic in the embedded clause in (11) with the obligatory positioning in the almost identical finite clause in (12), and the position of the clitic in (10) with that in (13).

(12) **=ɜj* *žon-ən* *=ɜj*
it.GEN know-PRS.1SG it.GEN
'I know it.'

[13] See for example Zaenen and Dalrymple (1996), Alsina (1997), and Butt (1997), within LFG, and Abeillé et al. (1998) within HPSG. The Romance data is discussed in detail by Monachesi (2005).

[14] Note that clitics are not hosted by C in this position, since there is no overt C in these embedded clauses. The clitic in (10) may also undergo SCC (to appear after *mɜn*).

(13) *=ɜm alə bon =ɜm kɜš-ən
it.ALL every day it.ALL look-PRS.3SG
'I watch it every day.'

Whatever formal analysis we adopt, in purely descriptive terms it remains the case that clitics cannot appear clause-initially in non-embedded clauses, but can in certain embedded clauses, in which case they are prosodically enclitic on the final word of the matrix clause. This suggests again that there are prosodic factors at work, ruling out absolute initial position for the clitic cluster in non-embedded clauses where there is no prosodic host to the left, but licensing absolute initial position in some embedded clauses.

The final constraint on clitic positioning in Ossetic is undoubtedly the most interesting, from both formal and typological perspectives. When a clause begins with a conjunction, any clitics in that clause are constrained to appear directly following the conjunction. Compare (14) with (5).

(14) ... ɜmɜ =dɜm [žawər-ə rɜšuʁd čənz] *=dɜm
and thee.ALL Zaur-GEN beautiful bride thee.ALL
ba-zərd-t-a
PV-speak-TR-PST.3SG
'... and the beautiful bride of Zaur called for you.'

Under either a parallel approach to conjunction, or a ConjP approach, this cannot be treated as second syntactic position within the CP, since the Conj does not appear in the CP. If the conjunction precedes a clause which contains a left-dislocated phrase or relative-correlative structure, the clitic cluster must appear *before* such phrases, that is outside the core CP.

It is therefore impossible to explain the position of the clitic cluster in examples such as (14) according to the same principles we have made use of in previous sections. At least in descriptive terms, it appears to be the case that the clitic cluster is 'attracted' out of its usual domain (the core CP) by a preceding conjunction. There are certain parallels between this possibility and what we have already seen: just as with weak clitic climbing, discussed in the previous section, the clitic cluster is able to appear at (or even to the left of) the left edge of its clause, as long as there is a prosodic host.

It is not only in the case of clausal conjunction that clitics are apparently 'attracted' to a preceding conjunction. When the first full phrase in a clause contains a conjunction, the clitic cluster may optionally appear directly following the embedded conjunction, even though this means that they appear inside a nominal constituent of which they are not themselves a part.

(15) [žawər ɜmɜ (=mɜm) alan] (=mɜm) ɜrba-səd-əštə
Zaur and me.ALL Alan me.ALL PV-go-PST.INTR.3PL
'Zaur and Alan came to me.'

There is no obvious prosodic or functional motivation for this (on the prosody, see below). The clitics concerned are not a part of the phrase in which they appear,

and there is simply no motivation for any position inside a nominal coordinate structure which could host clitics that do not even relate to that coordinate structure. Importantly, it does not matter how deeply embedded the conjunction is inside the first full phrase of the clause. In (16), the clitic appears following a conjunction embedded inside an adjectival phrase which is itself embedded inside the first full phrase of the clause.[15]

(16) *bɜχ, kʷətːɜr =əl a-bad-tɜn, afts =jɜ*
horse as.soon.as it.SUPER PV-sit-PST.1SG thus it.GEN
ba-mbɜršt-a, [[ɜdəχ ɜmɜ =jəl ɜnɜ-bon]$_{AdjP}$ lɜg]$_{NP}$ kɜj
PV-understand-PST.3SG weak and it.SUPER without-skill man that
bad-ə, wəj
sit-PRS.3SG that.GEN
'As soon as I mounted the horse, it immediately understood that a weak and unskillful man was riding it.' (ONC)

The relevant CP in this example runs from *ɜdəχ* to *bad-ə*; the first full phrase is the noun phrase *ɜdəχ ɜmɜ ɜnɜ-bon lɜg*, of which all but the final word is an embedded adjectival modifier. The clitic is placed not following the whole noun phrase, as we would expect, nor even following the adjective phrase, but inside the adjective phrase, directly following the conjunction. There is no way to explain the positioning of the clitic cluster in such examples by assuming that coordinated phrases can be split by processes such as topicalization. In Ossetic coordinated noun (or adjective) phrases cannot be split by any such processes.

This is the only context in which the clitic cluster may appear within a noun phrase; this context aside, there is a clear constraint against clitics appearing inside a noun phrase. For example, SCC is obligatory from case-marked infinitive phrases, which is most appropriately explained by assuming that case-marked infinitives do not head CPs or VPs, but noun phrases.

In this section we have seen roughly four distinct possibilities for clitic positioning in Ossetic, at least in purely descriptive terms. As a default, clitics appear in second syntactic position, i.e. following the first full phrase in their clause, or else following the first word, where this is a lexical or functional head (V or C). Clitics may also appear directly before the verb, when the only elements that precede the verb in the clause are proclitics. Alternatively, they may appear at the left edge of their clause in certain embedded clause types (weak clitic climbing). Finally, if a conjunction appears at (directly before) the left edge of a clause, then the clitics must directly follow the conjunction, and if there is no clausal conjunction but the first full phrase of the clause containts a conjunction, however deeply embedded, the clitics may appear directly after this conjunction. In the following section, we compare different strategies for dealing with this variety of positioning possibilities.

[15]It should be possible for the clitic to alternatively appear following the whole NP, as in (15), though only this form of the sentence exists in the corpus.

Let us summarize again the descriptive possibilities for clitic positioning, now with specific reference to the phrase-structure of Ossetic given above (1, 2).

(17)
a. Following first XP of core CP.
b. Following V, if V is first in CP.
c. Optionally, at left edge of embedded CP (WCC).
d. Optionally, between proclitics and V, if only proclitics precede the V.
e. At left edge of (outermost) CP, if clausal conjunction precedes the CP.
f. Optionally, directly following a phrasal conjunction that appears inside the first XP of the core CP.

4 Analysis

There are in principle a number of approaches that could be taken to the positioning of clitics in the Ossetic clause.

A number of recent treatments of clitics in LFG, e.g. by Bögel et al. (2010), Lowe (2011, 2014, 2015b) and Bögel (2015) involve languages with relatively free word order, and these analyses make use of the exocentric category S to model phrase-structure in their respective languages. This means that these analyses avoid the complications of hierarchical structure, and it is relatively simple, for example, to state that the clitic cluster appears in second syntactic position.

This may appear to be a promising approach to the clitic data. However, besides a default rule placing the CCL in second syntactic position, we would still need additional rules to license the appearance of clitics following an initial V or C, directly before V when two proclitics precede, at the left edge of an embedded S, at the left edge of an S preceded by a clausal conjunction, and the most problematic context: appearing inside the initial XP when this contains a phrasal conjunction. A lack of hierarchical structure would not necessarily make it any easier to account for any of these possibilities, and it certainly does not make it possible to reduce or collapse any of these possibilities by means of a more general statement. At the same time, we would have to abandon the generally agreed-upon hierarchical structure of the Ossetic clause, which is widely assumed on other grounds. Altogether, it does not seem that a flat analysis really gains us anything in accounting for clitic positioning, and it certainly loses something in the analysis of Ossetic c-structure more generally.

A further possibility, which at first sight appears promising but which ultimately does not live up to its promise, is the possibility of accounting for clitic positioning not by reference to syntax, but purely by reference to prosody. Purely prosodic approaches to clitic positioning in different languages are taken by authors such as Radanović-Kocić (1988, 1996), Anderson (1993, 1996, 2000, 2005), O'Connor (2002b) and Keydana (2011). However, all such accounts necessarily involve some interaction with syntax; even at the most basic level, the prosodic units with respect to which clitics are positioned *always* coincide with syntactic

units (e.g. positioning with reference to prosodic words, which always coincide with at least one full syntactic word). Significantly, there are no examples known to us that involve clitic positioning by reference to prosodic structures that specifically do not coincide with syntactic structure; for example, there are no examples of clitics that are invariably placed following the first syllable of a clause, regardless of whether that syllable coincides with a full word, or even a full morpheme. Since syntax must be involved, we believe that a fundamentally syntactic account should be the null hypothesis in any analysis of clitic positioning.

Nevertheless, prosodic positioning of some sort may seem a reasonable explanation for the Ossetic data. At the very least, based on the behaviour of WCC and preverbal proclitics, it is clear enough that prosody has some role to play in constraining clitic placement in Ossetic. But prosody by itself fails to provide an adequate account. As we have seen, the default position for clitics is following the first full XP of the clause. While initial XPs do generally carry a single strong stress in Ossetic, this is only the case for short and medium length XPs, i.e. those consisting of only a few words (which constitute the vast majority of examples). Very long XPs are also possible, e.g. noun phrases with embedded relative clauses, and it is not reasonable to assume that these carry a single stress or function as single prosodic units of equivalent status to those of short XPs. The generalization is therefore syntactic, not prosodic, even if many of the syntactic contexts share a similar prosody.

In addition, it is also not possible to explain purely by reference to prosody the positioning of clitics after a conjunction embedded inside the initial XP of a clause, and indeed in purely prosodic terms the similarity between this positioning and the obligatory positioning after a clausal conjunction is lost. Clausal conjunctions are characterized by stress, whereas phrasal conjunctions which can host clitics are stressless (and prosodically encliticized onto the preceding word). Thus the fact that conjunctions can host clitics, both when clausal and phrasal, cannot be explained by reference to prosody, but depends on their syntactic category.

We therefore see only one viable way of analysing clitic positioning in Ossetic, namely by reference to the hierarchical syntactic structure, but making use of additional strategies and constraints where these are required (and only where they are absolutely required). There is no one place in the hierarchical structure where the clitic cluster obviously adjoins in all examples; rather, it appears that the clitic cluster is almost unrestricted in terms of where it may adjoin (with the proviso that it is largely restricted to adjunction to clause-level categories). We therefore assume that the simplest approach is to permit the clitic cluster to (left-)adjoin almost anywhere in the c-structure, and then utilize other constraints to determine the correct adjunction site in any instance. The alternative would be to restrict the possible adjunction sites, even down to one site (e.g. adjunction to C, or absolute first position in CP), but then we would encounter more serious difficulties in accounting for the situations where the clitic cluster surfaces somewhere else (e.g. after the verb); to deal with this we might require recourse to concepts such as ‘movement’, however understood, in order to account for the surface position. We believe that

an approach which can do without 'movement' (whether prosodic or syntactic) is in principle to be preferred (though that is not to say that for some phenomena, in some languages, that may not be the best, or only, solution).

We therefore adopt an approach in which as many phrase-structure positions as possible are licensed for the clitic cluster, in order to permit as many of the attested surface positions as possible to be accounted for purely by reference to the phrasal syntax. One of the difficulties with any hierarchical account is how to deal with the fact that the clitic cluster can appear following in principle any XP position that precedes the verb, as long as it is the first filled XP position in the CP. That is, considering again the trees in (1) and (2), there are a variety of XP positions within the core CP, all of which are optional, and all of which could therefore potentially serve as the first XP in the clause. The phrase in Spec,CP is a default topic, while elements in Spec,VP and adjoined to VP function as foci. But whether the first XP in a clause is topical or focal, the clitic cluster will directly follow it. One way to deal with this is to permit the clitic cluster node, the CCL, to adjoin to the left of any clausal head or projection, via the following rules:

(18)

a.	CP	→	CCL	CP	d.	V′	→	CCL	V′
			↑=↓	↑=↓				↑=↓	↑=↓
b.	C′	→	CCL	C′	e.	S	→	CCL	S
			↑=↓	↑=↓				↑=↓	↑=↓
c.	VP	→	CCL	VP	f.	V	→	CCL,	V
			↑=↓	↑=↓				↑=↓	↑=↓

The rules in (18a–e) license adjunction of CCL to the left of CP, C′, VP, V′ and S. The rule in (18f) licenses adjunction of CCL to the left or right of V; this variability in positioning is required to account for orderings like that in (8), where the CCL directly follows the verb. Together, these rules essentially amount to the statement that clitics may appear anywhere in the clause, up to the immediately post-verbal position. Clearly, therefore, the application of these rules must be constrained in such a way that the CCL only ever actually appears following the initial XP, or the verb if it is clause-initial. To this end, we must assume a general constraint (formalized below) to force the clitic cluster to appear as far leftwards in the clause as possible. For example, if we have both a topical XP in Spec,CP and a focal XP in Spec,VP, in principle the clitic cluster could be adjoined to either C′ or V′, but in fact it must appear following the leftmost, topical, XP. WCC and the attraction of clitics to clause-preceding coordinating conjunctions can be handled in a similar way as adjunction to the uppermost CP.

This rule can also account for the appearance of clitics between the proclitic verbal particles and the V, inside the verbal complex. However, in conjunction with the rule requiring clitics to appear leftmost in their clause, it is then problematic to account for the appearance of the clitic cluster directly before the verb when more than one proclitic precedes and when the first of those proclitics can be analysed as the head of a separate, full, XP. This is accounted for below.

In addition to the rules in (18), we also require a rule licensing the appearance

of a CCL as a daughter of the exocentric category S. It is possible that the first and second XPs in a clause are both XP daughters of S, and if this is so the CCL must appear following the first XP, meaning that it too must be a daughter of S. In order to capture the fact that, within S, the CCL must appear following the first XP, whatever that is, we must assume the following rule expanding S:[16]

(19) S → (XP CCL) XP* VP (CCL) XP*
(↑ GF) =↓ ↑=↓ (↑ GF) =↓ ↑=↓ ↑=↓ (↑ GF) =↓

Finally, we must account for the positioning of clitics after an embedded phrasal conjunction. Since we cannot say how embedded the Conj is, it is difficult, but not impossible, to state a functional constraint on the CCL that will permit it to 'reach out' of the phrase within which it is embedded and make its contribution at the clausal level, and that would not also permit it to make a contribution at a higher level, if it were part of an embedded clause.

(20) Conj → Conj CCL
↑=↓ ((∈)GF* ↑) =↓
¬(→ TENSE)

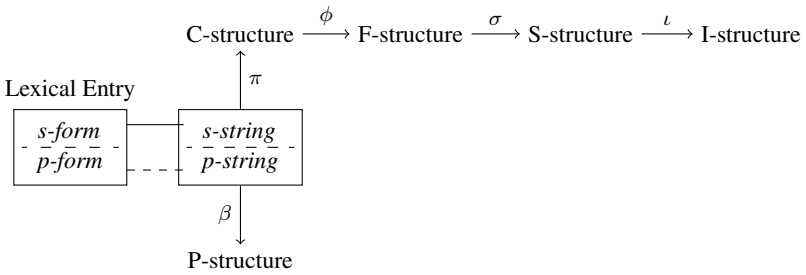
As mentioned, in addition to the ordinary phrase-structure rules licensing adjunction of the CCL, we must also assume a number of more general constraints which serve to restrict the adjunction possibilities, such that there are at most two, and usually only one, possible positions for the CCL in any instance. We propose to model this using OT constraints, with stochastic re-ranking to account for cases where two positions are possible. In this, we build on the OT-LFG analysis of clitic positioning proposed by Lowe (2015a), although the constraints we require are somewhat different.

In Optimality Theoretic approaches to LFG, any full or partial structure (usually an f-structure) serves as the INPUT, a 'universal' LFG grammar is the GEN, and the set of candidates produced by applying the INPUT to GEN are evaluated according to a set of ranked constraints EVAL (on Optimality Theoretic LFG see, for example, Bresnan 1996, 2000a,b; Frank et al. 1998; Kuhn 1999, 2001). While early Optimality Theoretic approaches to LFG focused primarily on the core projections, f-structure and c-structure, Lowe (2015a) proposes a more complex system, based on a more complex projection architecture. The constraints against which candidates are evaluated may refer to any structure in the projection architecture, or to any projection function between structures. For Bresnan (2000a: 15), each candidate was a quadruple consisting of a c-structure, an f-structure, the mapping from the lexicon to c-structure and the mapping from c-structure to f-structure. For Lowe (2015a), the candidates are fifteen-fold, based on the architecture shown in (21) (Dalrymple and Mycock 2011, Mycock and Lowe 2013):[17]

[16] Another option would be to abandon S, replacing it by an endocentric category, perhaps I, meaning that rules parallel to those in (18) could be stated. But there is no evidence that any element heads I in Ossetic.

[17] The fifteen aspects are: s–string, p–string, p–structure, c–structure, f–structure, s–structure, i–

(21)



A frequent criticism of OT is that an infinite set of candidates makes the grammar too unconstrained and computationally intractable. This is a basic theoretical principle of OT and is thus carried over to those variants of OT-LFG which strive to override existing LFG mechanisms. For example, Bresnan (2000a) envisages OT-LFG as a general model of grammar that completely replaces the traditional c-structure rules and f-structure equations. Such an approach does require an infinite set of candidates to function properly, as there is no other way of accounting for grammaticality and cross-linguistic variation except for inherently violable OT constraints.

However, most actual applications of OT to LFG, e.g. Butt et al. (1997), Broadwell (1999), are not as radical. They only analyze competition between those structures which are difficult to disambiguate using the standard, inviolable, grammatical constraints. The role of OT is thus merely that of an additional filtering mechanism that models competition between several possible realizations of a single underlying structure. In such applications of OT, it is commonly assumed that the candidate set is generated by a normal LFG grammar. This makes it finite by definition, therefore rendering the aforementioned criticism invalid. Since all candidates involved in the computation are generated by a normal LFG grammar, such an approach is not any more unconstrained than standard LFG.

At the same time, it cannot be disputed that some sort of OT-like filtering mechanism must be present in any grammatical theory for resolving syntactic ambiguity, whose range in natural language is far greater than any traditional rule-based formalism can handle. This is the motivation behind the use of OT-like constraints in XLE (Frank et al. 1998).[18] If an OT-like mechanism is required for disambiguation in any case, extending its use for other purposes ought not to significantly increase computational complexity.

For our purposes, the relevant constraints refer either to the s(yntactic)-string or the p(rosodic)-string, or to the mapping between the string and c-structure (π). We assume the following constraints:[19]

structure, β, π, ϕ, σ, ι, two 'lexicalization functions', which map from the s–string and p–string to the (language-specific) lexicon, and the interface between s–string and p–string.

[18]The only difference is that XLE OT-like constraints only evaluate potential analyses of a given sentence but do not examine other potential candidates with the same f-structure. Therefore, any sentence that is allowed by the LFG grammar will be considered grammatical, even if it is ruled out by competition with other candidates.

[19]Some of these are taken from Lowe (2015a), partially building on proposals for OT constraints

(22)
a. ALIGN(CL,L;PW,R): align the left edge of a prosodic clitic with the right edge of a prosodic word. (A prosodic constraint preventing enclitics from appearing at left edge of prosodic unit.)

b. *MOVE–RT: Do not move a clitic rightwards in the π mapping. (Constraint penalizing 'movement' of clitics.)

c. *MOVE–LFT: Do not move a clitic leftwards in the π mapping. (Constraint penalizing 'movement' of clitics (prosodic inversion).)

d. *CL-IN-NP: No word that appears at the left edge of an NP may appear to the left of a clitic, unless the word that appears at the right edge of that NP also appears to the left edge of the clitic. (Constraint against clitics appearing inside NP.)

e. *ALIGN(CL,L;DOM,L): Do not align the left edge of a clitic with the left edge of its syntactic domain. (Syntactic constraint against clitics at left edge of CP.)

f. ALIGN(CL,L;DOM,L): align the left edge of a clitic with the left edge of its clause. (Syntactic constraint constraining clitic to appear as far leftward in the clause as possible.)[20]

g. *CL-BTW-PROCL: Penalize every enclitic that appears between two proclitics. (Prosodic constraint against (en)clitics appearing between two proclitics, i.e. prefer [Pro Pro$^+$ Enc] over [Pro$^+$ Enc Pro$^+$].)[21]

We assume the following ranking, with stochastic range as indicated. The competition between constraints f and g licenses the optional positioning listed in (17d) above; competition between f and e licenses the optionality listed in (17c); competition between f and d licenses the optionality in (17f). Constraints a, b and c need never be violated in accounting for Ossetic clitics, but they are included since they rule out theoretically possible structures, such as clitics first in a standalone clause, or analyses involving 'movement'.[22]

(23) Constraint ranking, showing stochastic range:
$[_a$ a $]_a$ > $[_b[_c$ b,c $]_b]_c$ > $[_d[_e$ d,e > $[_f$ f $]_d]_e$ $[_g$ $]_f$ g $]_g$

5 Examples

When clitics appear in second syntactic position, whether this is following the first clause-level XP, as in (4), (5), (6), and (7), or following an initial V or C, as in (8), or at the start of a clause following a clausal conjunction (14), the PS-rules proposed above, in combination with constraint f, requiring clitics to appear as far

on clitics by Legendre (e.g. 2000), Grimshaw (2001), and Woolford (2002).

[20]Following McCarthy and Prince (1993), alignment constraints can be multiply violated, so the further left a clitic appears, the fewer violations it incurs.

[21]This constraint may appear somewhat ad hoc, but seems intuitively plausible, and fits with patterns of proclitic-enclitic alignment in related languages such as Ancient Greek and Vedic Sanskrit (Lowe 2014).

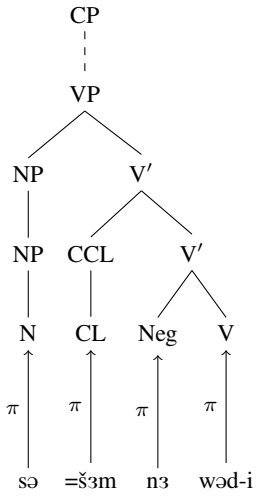
[22]Constraints b and c are treated as separate constraints following Lowe (2015a); Lowe (2015a) ranks b above c, but this is not relevant to the present analysis.

left as possible, will unproblematically account for the position of the clitic cluster. We omit exemplifying this here.

The variable positioning of clitics when the clause begins with more than one proclitic immediately preceding the verb, as seen in (9), is captured by the variability in ranking of constraints f and g. This is illustrated in (25) and (26), which provide analyses for (9=24).[23]

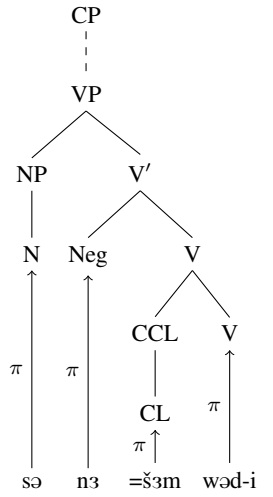
(24) *sə* *(=šɜm)* *nɜ* *(=šɜm)* *wəd-i* **=šɜm,* *aχɜm* *nɜ-j*
what they.ALL NEG they.ALL be-PST.INTR.3SG they.ALL such NEG-is
'There is nothing that they did not have (lit. that to them there was not).' (spoken text)

(25)



Input = f-structure		a	b	f	g
a.	*c–str.* =š. *s.* *n.* *w.* *s–str.* =š. *s.* *n.* *w.* *p–str.* =š. *s.* *n.* *w.*	*!			
☞ b.	*c–str.* *s.* =š. *n.* *w.* *s–str.* *s.* =š. *n.* *w.* *p–str.* *s.* =š. *n.* *w.*			*	*
c.	*c–str.* *s.* *n.* =š. *w.* *s–str.* *s.* *n.* =š. *w.* *p–str.* *s.* *n.* =š. *w.*			**!	
d.	*c–str.* *s.* *n.* =š. *w.* *s–str.* *s.* =š. *n.* *w.* *p–str.* *s.* =š. *n.* *w.*		*!	*	

(26)



Input = f-structure		a	b	g	f
a.	*c–str.* =š. *s.* *n.* *w.* *s–str.* =š. *s.* *n.* *w.* *p–str.* =š. *s.* *n.* *w.*	*!			
b.	*c–str.* *s.* =š. *n.* *w.* *s–str.* *s.* =š. *n.* *w.* *p–str.* *s.* =š. *n.* *w.*			*!	*
☞ c.	*c–str.* *s.* *n.* =š. *w.* *s–str.* *s.* *n.* =š. *w.* *p–str.* *s.* *n.* =š. *w.*				**
d.	*c–str.* *s.* *n.* =š. *w.* *s–str.* *s.* =š. *n.* *w.* *p–str.* *s.* =š. *n.* *w.*		*!		*

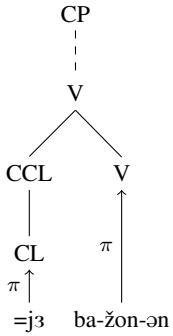
The optionality of 'Weak Clitic Climbing' is captured by the re-ranking of constraints e and f. This is exemplified in (28) and (29), for the sentence in (11=27). For the optional alignment of the clitic cluster following a phrasal conjunction embedded inside the first XP of a clause, the two possibilities are obtained by the

[23]Here we illustrate merely the potential effects of varying the stochastic weights of the constraints; a greater volume of naturally occurring data would be required to make claims as to the actual weights.

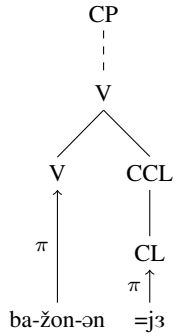
variable ranking of constraints d and f. This is exemplified in (31) and (32), on the following page, for the sentence in (15=30).

(27) *a-sə* *tɜkːɜ* *šaχat-əl* *=mɜ* *qɜw-ə* *[(=jɜ)*
this-ATTR exactly hour-SUPER me.GEN need-PRS.3SG it.GEN
ba-žon-ən *(=ɜj)]*
PV-know-INF it.GEN
'It is exactly now that I need to know it.' (ONC: *Max dug* 9, 2007)

(28)



(29)



Input = f-structure		a	b	c	f	e
☞ a.	*c–str. =jɜ b.* *s–str. =jɜ b.* *p–str. =jɜ b.*					*
b.	*c–str. b. =ɜj* *s–str. b. =ɜj* *p–str. b. =ɜj*				*!	
c.	*c–str. b. =ɜj* *s–str. =jɜ b.* *p–str. =jɜ b.*		*!			*
d.	*c–str. =jɜ b.* *s–str. b. =ɜj* *p–str. b. =ɜj*			*!	*	

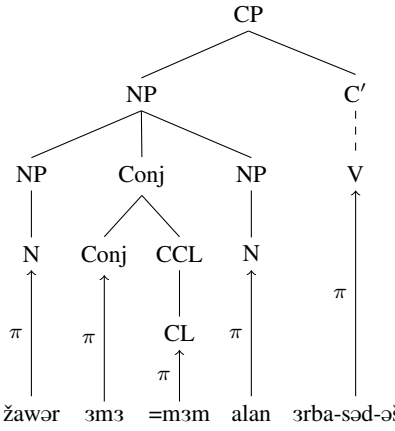
Input = f-structure		a	b	c	e	f
a.	*c–str. =jɜ b.* *s–str. =jɜ b.* *p–str. =jɜ b.*				*!	
☞ b.	*c–str. b. =ɜj* *s–str. b. =ɜj* *p–str. b. =ɜj*					*
c.	*c–str. b. =ɜj* *s–str. =jɜ b.* *p–str. =jɜ b.*		*!		*	
d.	*c–str. =jɜ b.* *s–str. b. =ɜj* *p–str. b. =ɜj*			*!		*

6 Conclusion

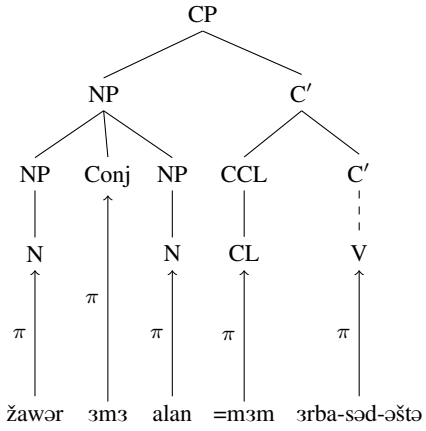
The data on clitic positioning in Ossetic is complex, suggesting that a number of interacting factors are at work in constraining clitic position. Some of these factors are clearly syntatic, others appear to be prosodic. We have shown that this can be modelled in LFG using a stochastic version of OT which permits constraints from different components of the grammar, e.g. syntax and prosody, to jointly contribute to the evaluation of the optimal candidate. Our analysis does not require clitic 'movement', and permits us to capture the variable positioning of the Ossetic clitic cluster without compromising the widely-agreed hierarchical structure of the Ossetic clause. Our analysis involves almost unrestricted adjunction possibilities for the Ossetic clitic cluster, constrained by OT constraints, and this enables us to capture the linear aspect of clitic positioning without abandoning the hierarchical c-structure.

(30) *[žawər ɜmɜ (=mɜm) alan] (=mɜm) ɜrba-səd-əštə*
Zaur and me.ALL Alan me.ALL PV-go-PST.INTR.3PL
'Zaur and Alan came to me.'

(31)



(32)



Input = f-structure		c	f	d
a.	*c–str. ž. ɜm. a. =m. ɜr.* *s–str. ž. ɜm. a. =m. ɜr.* *p–str. ž. ɜm. a. =m. ɜr.*		**!*	
☞ b.	*c–str. ž. ɜm. =m. a. ɜr.* *s–str. ž. ɜm. =m. a. ɜr.* *p–str. ž. ɜm. =m. a. ɜr.*		**	*
c.	*c–str. =m. ž. ɜm. a. ɜr.* *s–str. ž. =m. ɜm. a. ɜr.* *p–str. ž. =m. ɜm. a. ɜr.*	*!	*	*

Input = f-structure		c	d	f
☞ a.	*c–str. ž. ɜm. a. =m. ɜr.* *s–str. ž. ɜm. a. =m. ɜr.* *p–str. ž. ɜm. a. =m. ɜr.*			***
b.	*c–str. ž. ɜm. =m. a. ɜr.* *s–str. ž. ɜm. =m. a. ɜr.* *p–str. ž. ɜm. =m. a. ɜr.*		*!	**
c.	*c–str. =m. ž. ɜm. a. ɜr.* *s–str. ž. =m. ɜm. a. ɜr.* *p–str. ž. =m. ɜm. a. ɜr.*	*!	*	*

References

Abeillé, Anne, Godard, Danièle and Sag, Ivan A. 1998. Two kinds of composition in French complex predicates. In Erhard Hinrichs, Andreas Kathol and Tsuneko Nakazawa (eds.), *Complex Predicates in Nonderivational Syntax*, pages 1–41, New York: Academic Press.

Alsina, Alex. 1997. A theory of complex predicates: evidence from causatives in Bantu and Romance. In Alex Alsina, Joan Bresnan and Peter Sells (eds.), *Complex Predicates*, pages 203–246, Stanford, CA: CSLI Publications.

Anderson, Stephen R. 1993. Wackernagel's Revenge: Clitics, Morphology, and the Syntax of Second Position. *Language* 69(1), 68–98.

Anderson, Stephen R. 1996. How to Put Your Clitics in Their Place. *Linguistic Review* 13, 165–191.

Anderson, Stephen R. 2000. Towards an Optimal Account of Second-Position Phenomena. In Joost Dekkers, Frank van der Leeuw and Jeroen van de Weijer (eds.), *Optimality Theory: Phonology, Syntax and Acquisition*, pages 302–333, Oxford: Oxford University Press.

Anderson, Stephen R. 2005. *Aspects of the Theory of Clitics*. Oxford: Oxford University Press.

Austin, Peter and Bresnan, Joan. 1996. Non-Configurationality in Australian Aboriginal Languages. *Natural Language and Linguistic Theory* 14(2), 215–268.

Belyaev, Oleg I. and Haug, Dag T. T. 2014. Pronominal Coreference in Ossetic Correlatives and the Syntax-Semantics Interface. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG14 Conference*, pages 89–109, Stanford, CA: CSLI Publications.

Bögel, Tina. 2010. Pashto (Endo-)Clitics in a Parallel Architecture. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG10 Conference*, pages 85–105, Stanford, CA: CSLI Publications.

Bögel, Tina. 2014. Degema and the String Interface. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG14 Conference*, pages 110–130, Stanford, CA: CSLI Publications.

Bögel, Tina. 2015. *The Syntax-Prosody Interface in Lexical Functional Grammar*. Ph. D.thesis, Universität Konstanz.

Bögel, Tina, Butt, Miriam, Kaplan, Ronald M., King, Tracy Holloway and Maxwell III, John T. 2010. Second Position and the Prosody-Syntax Interface. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG10 Conference*, pages 106–126, Stanford, CA: CSLI Publications.

Bresnan, Joan. 1996. LFG in an OT Setting: Modelling Competition and Economy. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG96 Conference*, Stanford, CA: CSLI Publications.

Bresnan, Joan. 2000a. Optimal Syntax. In Joost Dekkers, Frank van der Leeuw and Jeroen van de Weijer (eds.), *Optimality Theory: Phonology, Syntax and Acquisition*, pages 334–385, Oxford: Oxford University Press.

Bresnan, Joan. 2000b. Explaining Morphosyntactic Competition. In Mark Baltin and Chris Collins (eds.), *Handbook of Contemporary Syntactic Theory*, Oxford: Blackwell.

Broadwell, George Aaron. 1999. The interaction of focus and constituent order in San Dionicio Zapotec. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG99 Conference*, Stanford, CA: CSLI Publications.

Butt, Miriam. 1997. Complex Predicates in Urdu. In Alex Alsina, Joan Bresnan and Peter Sells (eds.), *Complex Predicates*, Stanford, CA: CSLI Publications.

Butt, Miriam, Dalrymple, Mary and Frank, Annette. 1997. An architecture for linking theory in LFG. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG97 Conference*, Stanford, CA: CSLI Publications.

Dalrymple, Mary and Mycock, Louise. 2011. The Prosody-Semantics Interface. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG11 Conference*, pages 173–193, Stanford, CA: CSLI Publications.

Dalrymple, Mary and Nikolaeva, Irina. 2011. *Objects and Information Structure*. Cambridge: Cambridge University Press.

Dzaxova, Veronika T. 2009. *Foneticeskie xarakteristiki fonologiceskoj sistemy sovremennogo osetinskogo (ironskogo) literaturnogo jazyka [The phonetic characterisics of the phonological system of contemporary literary Iron Ossetic]*. Vladikavkaz: Izdatel'stvo SOGPI.

Erschler, David. 2010. Modularity and 2P clitics: Arguments from Digor Ossetic. In Yehuda N. Falk (ed.), *Online Proceedings of the twenty-fifth annual conference of the Israel Association for Theoretical Linguistics (IATL 25)*.

Erschler, David. 2012. From preverbal focus to preverbal "left periphery": The Ossetic clause architecture in areal and diachronic perspective. *Lingua* 122, 673–699.

Frank, Anette, King, Tracy Holloway, Kuhn, Jonas and Maxwell, John. 1998. Optimality Theory Style Constraint Ranking in Large-scale LFG Grammars. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG98 Conference*, Stanford, CA: CSLI Publications.

Franks, Steven and King, Tracy Holloway. 2000. *A Handbook of Slavic Clitics*. Oxford: Oxford University Press.

Grimshaw, Jane. 2001. Optimal Clitic Position and the Lexicon in Romance Clitic Systems. In Géraldine Legendre, Jane Grimshaw and Sten Vikner (eds.), *Optimality-Theoretic Syntax*, pages 205–240, Cambridge, MA: MIT Press.

Keydana, Götz. 2011. Wackernagel in the Language of the Rigveda. *Historische Sprachforschung* 124, 80–107.

Kuhn, Jonas. 1999. Two ways of formalizing OT Syntax in the LFG framework. MS, Institut für maschinelle Sprachverarbeitung, Universität Stuttgart. Downloaded from `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.34.747` November 2011.

Kuhn, Jonas. 2001. Generation and Parsing in Optimality Theoretic Syntax: Issues in the Formalization of OT-LFG. In Peter Sells (ed.), *Formal and Empirical Issues in Optimality Theoretic Syntax*, pages 313–366, Stanford, CA: CSLI Publications.

Legendre, Géraldine. 2000. Positioning Romanian Verbal Clitics at PF: An Optimality Theoretic Analysis. In Birgit Gerlach and Janet Grijzenhout (eds.), *Clitics in Phonology, Morphology and Syntax*, pages 219–254, Amsterdam: Benjamins.

Lowe, John J. 2011. Ṛgvedic Clitics and 'Prosodic Movement'. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG11 Conference*, pages 360–380, Stanford, CA: CSLI Publications.

Lowe, John J. 2014. Accented clitics in the *Ṛgveda*. *Transactions of the Philological Society* 112(1), 5–43.

Lowe, John J. 2015a. Clitics: Separating Syntax and Prosody. *Journal of Linguistics* FirstView article available on CJO, DOI: 10.1017/S002222671500002X. 45pp. Currently scheduled to appear in vol. 52(2), July 2016.

Lowe, John J. 2015b. *Participles in Rigvedic Sanskrit: The Syntax and Semantics of Adjectival Verb Forms*. Oxford: Oxford University Press.

Lyutikova, Ekaterina and Tatevosov, Sergei. 2009. The clause internal left edge: Exploring the preverbal position in Ossetian. Paper presented at the Third International Conference on Iranian Languages. Université Paris 3 – Sorbonne Nouvelle. http://otipl.philol.msu.ru/staff/people/tatevosov/clause_internal_left_edge_4.pdf.

McCarthy, John J. and Prince, Alan S. 1993. Generalized Alignment. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1993*, pages 79–153, Dordrecht: Foris.

Monachesi, Paola. 2005. *The Verbal Complex in Romance: A Case Study in Grammatical Interfaces*. Oxford: Oxford University Press.

Mycock, Louise and Lowe, John J. 2013. The Prosodic Marking of Discourse Functions. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG13 Conference*, pages 440–460, Stanford: CSLI Publications.

Newman, Paula S. 1996. Representing 2P Clitic Placement in LFG. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG96 Conference*, Stanford, CA: CSLI Publications.

Nordlinger, Rachel. 1998a. *A Grammar of Wambaya, Northern Australia*. Canberra: Pacific Linguistics.

Nordlinger, Rachel. 1998b. *Constructive Case: Evidence from Australian Languages*. Stanford, CA: CSLI Publications.

O'Connor, Rob. 2002a. Clitics and Phrasal Affixation in Constructive Morphology. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG02 Conference*, pages 315–332, Stanford, CA: CSLI Publications.

O'Connor, Rob. 2002b. The placement of enclitics in Bosnian, Croatian and Serbian. MS, University of Manchester. Available at http://roa.rutgers.edu/files/521-0502/521-0502-OCONNOR-0-0.PDF.

Radanović-Kocić, Vesna. 1988. *The Grammar of Serbo-Croatian Clitics: A Synchronic and Diachronic Perspective*. Ph. D.thesis, University of Illinois at Urbana-Champaign.

Radanović-Kocić, Vesna. 1996. The Placement of Serbo-Croatian Clitics: A Prosodic Approach. In Aaron L. Halpern and Arnold M. Zwicky (eds.), *Approaching Second: Second Position Clitics and Related Phenomena*, pages 429–445, Stanford, CA: CSLI Publications.

Wescoat, Michael Thomas. 2009. Udi Person Markers and Lexical Integrity. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG09 Conference*, pages 604–622, Stanford, CA: CSLI Publications.

Woolford, Ellen. 2002. Clitics and agreement in competition: ergative cross-referencing patterns. In A. Carpenter, A. Coetzee and De Lacy P. (eds.), *Papers in Optimality Theory II*, pages 421–449, University of Massachusetts, Amherst: GLSA.

Zaenen, Annie and Dalrymple, Mary. 1996. Les Verbes Causatifs "Polymorphiques": Les Prédicats Complexes en Français. *Langages* 122, 79–95.

Zwicky, Arnold M. 1977. *On Clitics*. Bloomington: Indiana University Linguistics Club.

GENERALISING FUNCTIONAL CATEGORIES IN LFG

Ryo Otoguro
Waseda University

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King (Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

This papers aims to present a lexicalist analysis of verb placement variation found in Germanic and Romance languages. Unlike derivational approaches to syntax, Lexical Functional Grammar accounts for different phrase structure positions of finite verbs by base-generating them in functional heads. The present study proposes a framework that allows us to formulate the conditions that license finite verbs to bear a functional category status. Crucial in the analysis is the encoding patterns of person features in lexical entries and their paradigmatic organisation of the lexicon. I will further show that the current proposal can be supported by dialectal variation and diachronic change of verb placement.

1 Introduction

In most Germanic languages, finite verbs are placed in the second position (V2) in main clauses. (1) exemplifies the V2 structure in Yiddish, Icelandic, Danish and Swedish, in which the subject appears in the clause initial position and the finite verb follows it.

(1) a. Max shikt nit avek dem brif.
M. sends not away the letter
'Max doesn't mail the letter.' (Yiddish)

b. Jón keypti ekki bókina.
J. bought not the.book
'John didn't buy the book.' (Icelandic)

c. Peter drikker ofte kaffe.
P. drinks often coffee
'Peter often drinks coffee.' (Danish)

d. Johan köpte inte boken.
J. bought not the.book
'John didn't buy the book.' (Swedish) (cf. Rohrbacher, 1999, 12)

The clause initial position can be filled by a non-subject constituent in V2 clauses. When a non-subject constituent is placed in the clause initial position, the finite verb precedes the subject as shown in (2).

(2) a. Dos bukh shik ikh avek.
the book send I away
'I am mailing the book.' (Yiddish)

[†]I would like to thank Ash Asudeh, Mary Dalrymple, Louisa Sadler and Andrew Spencer for discussion. I am also indebted to two anonymous reviewers for their helpful comments and stylistic suggestions. This work is supported by Grant-in-Aid for Scientific Research (C), Japan Society for the Promotion of Science (Grant No.26370579).

b. Dette spørsmålet skjønte Jens ikke.
this question understood J. not
'This question John didn't understand.' (Norwegian)

c. Igår köpte Lena en ny bok.
yesterday bought L. a new book
'Yesterday Lena bought a new book.' (Swedish)

(Rohrbacher, 1999, 12–3)

In embedded clauses, on the other hand, Germanic languages exhibit variation with regard to the location of a finite verb. The clause-initial position is filled by the subject in all the languages under discussion, but the relative order between a finite verb and an adverb varies. In Icelandic and Yiddish, a finite verb immediately follows the subject and precedes an adverb as in (3). In Scandinavian languages and Faroese, on the other hand, a finite verb follows an adverb as in (4).[1]

(3) a. ... að Jón borðar oft tómata.
that J. eats often tomatoes
'... that John often eats tomatoes.' (Icelandic)

b. *... að Jón oft borðar tómata. (Icelandic)

c. ... az Jonas est oft pomidorn.
that J. eats often tomatoes (Yiddish)

d. *... az Jonas oft est pomidorn. (Yiddish)

(4) a. ... at Johan ofte spiser tomater.
that J. often eats tomatoes (Danish)

b. *... at Johan spiser ofte tomater. (Danish)

c. ... at Jón ofta etur tomatir.
that J. often eats tomatoes (Faroese)

d. *... at Jón etur ofta tomatir. (Faroese) (Vikner, 1997, 189)

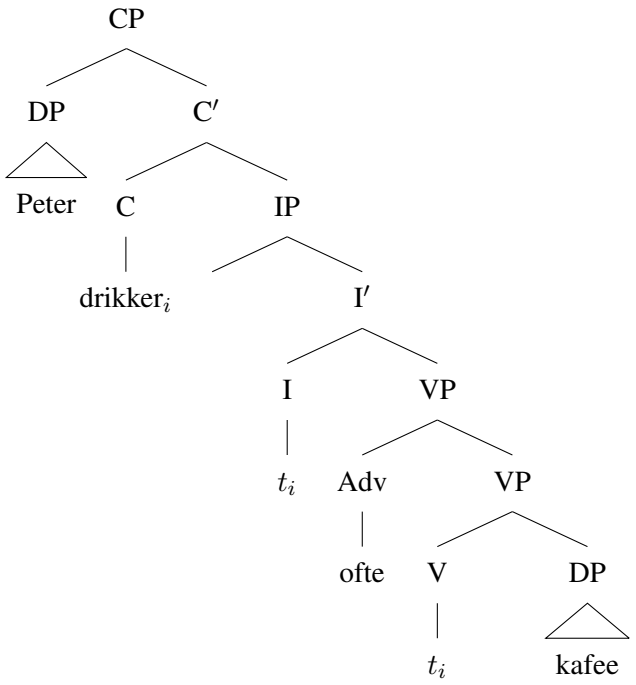
2 Capturing verb position variation

To capture the fact that a finite verb stays in the second position in main clauses and the clause initial position is occupied by various types of constituent, it is normally assumed in a derivational approach to syntax that the finite verb undergoes movement operations from V-to-I and I-to-C and lands in C position. Under this assumption, the clause-initial element is placed in Spec,CP preceding the finite verb as illustrated in (5) (cf. Holmberg and Platzack, 1995; Vikner, 1995).[2]

[1] Faroese exhibits some exceptions. See section 5 for details.

[2] Travis (1991), Sells (2001) and Toivonen (2003) propose a different structure for subject initial V2 clauses, in which the finite verb is in I, not C, and the subject is in Spec,IP. Also, for ease of exposition, Spec,VP is omitted.

(5)



With regard to the embedded clause ordering, the relative order between a finite verb and an adverb in Icelandic and Yiddish is the same as that of French and Italian main clauses ((6) and (7)). The verb placement in languages like French and Italian is normally accounted for by postulating V-to-I movement in generative derivational analyses (cf. Emonds, 1978; Pollock, 1989), so by adopting the same assumption it has been proposed that finite verbs in Icelandic and Yiddish embedded clauses also undergo V-to-I movement. The reason that the verb does not move up to C in an embedded clause is attributed to the presence of a complementiser that occupies the C position, that is the complementiser blocks the verb from moving further up to C, as shown in (8).

(6) a. Jean embrasse souvent Marie.
J. kisses often M.
'John often kisses Mary.'

b. *Jean souvent embrasse Marie.
J. often kisses M.

c. Jean (ne) mange pas de chocolat.
J. (NEG) eats not of chocolate
'John doesn't eat chocolate.'

d. *Jean (ne) pas mange de chocolat.
J. (NEG) not eats of chocolate (French)

(Pollock, 1989, 367)

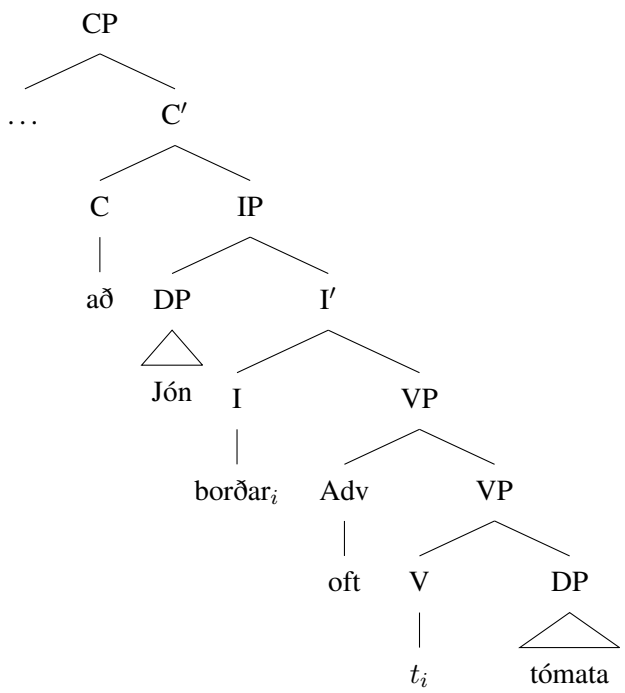
(7) a. Quel medico risolverà completamente i tuoi problemi.
that doctor solve.FUT completely the your problems
'That doctor will solve your problem completely.'

b. Quel medico risolverà i tuoi problemi *completamente*.
that doctor solve.FUT the your problem completely

c. *Quel medico completamente risolverà i tuoi problemi.
that doctor completely solve.FUT the your problem (Italian)

(Rohrbacher, 1999, 209)

(8)



In contrast, the relative order between the finite verbs and the adverbs in Scandinavian languages and Faroese in (4) suggests that those languages do not exhibit V-to-I movement. That is, unlike Icelandic and Yiddish, a finite verb follows an adverb in embedded clauses in those languages. Therefore, the question arises as to why those two types of Germanic languages place a finite verb in different positions in embedded clauses.

To resolve this puzzle, it has been pointed out that one of the notable differences between the two types of languages is the degree of richness of their verb agreement morphology. Table 1 illustrates the contrast between the verb inflectional paradigms in Yiddish, Icelandic and Danish. Yiddish and Icelandic clearly exhibit more morphological distinctions across different person number combinations, while Danish does not show any agreement morphology.

Some attempts have been made to relate rich agreement morphology and the presence of V-to-I movement. One of them can be stated as the so-called Rich

	Yiddish *loyf-n* 'run'		Icelandic *segj-a* 'say'		Danish *høre* 'heard'	
	Sg	Pl	Sg	Pl	Sg	Pl
1	loyf	loyf-n	segi	segj-um	hør-te	hør-te
2	loyf-st	loyf-t	segi-r	seg-ið	hør-te	hør-te
3	loyf-t	loyf-n	segi-r	segj-a	hør-te	hør-te

Table 1: Yiddish, Icelandic and Danish verb paradigms

Agreement Hypothesis as in (9).[3]

(9) Rich Agreement Hypothesis:
Rich agreement morphology induces V-to-I movement
(Kosmeijer, 1986; Rohrbacher, 1999; Koeneman and Zeijlstra, 2010, 2012)

A more specific formulation is given by Rohrbacher (1999) as follows:

(10) The Paradigm-Verb Raising Correlate:
A language has V to I raising if and only if in at least one number of one tense of the regular verb paradigm(s), the person features [1ST] and [2ND] are both distinctively marked. (Rohrbacher, 1999, 116)

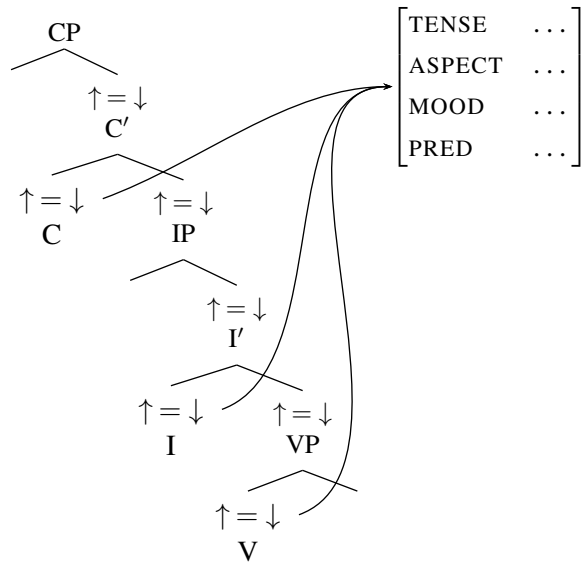
When a language satisfies the person feature distinctions stated in (10), that is the first person and second person forms are in distinct morphological shapes at least in one number, e.g., singular or plural, of one tense, e.g., past or present, I is qualified as what Rohrbacher calls a referential category. Under this assumption, lexically listed inflectional affixes are realised in referential I and a verb moves from V to I to form a fully inflected form there (Rohrbacher, 1999, 130).

Capturing variation in verb placement is also an issue for non-derivational frameworks. In Lexical Functional Grammar (LFG), c(onstituent)-structure is a purely surface representation of hierarchical relations and linear ordering of words and phrases, so all lexical items are base-generated and no derivational operation including movement is postulated. Hence, the category of finite verbs is often specified as I or C in their lexical entry, so that they can appear in a higher surface position in phrase structure. This approach has been attested in languages like Tagalog (Kroeger, 1993), Russian (King, 1995), Welsh (Sadler, 1997), Irish (Asudeh, 2012), Swedish (Sells, 2001; Toivonen, 2003), Icelandic (Sells, 2003, 2005), European Portuguese (Luís and Otoguro, 2004, 2005, 2011) and many others. Since functional heads are mapped onto the same level of f(unctional)-structure as their lexical head (Bresnan et al., 2016, 105-7), TENSE, ASPECT, MOOD and PRED features as well as agreement features specified in the verb's lexical entry are contributed to the same f-structure regardless of whether the verb is located in V, I or

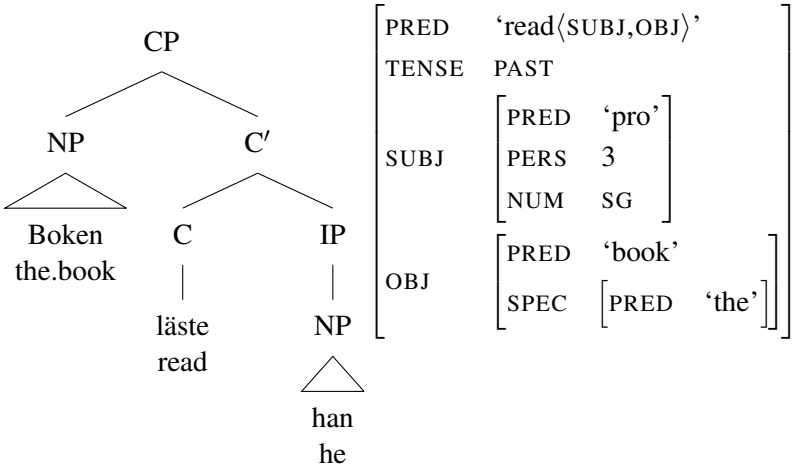
[3]Bobaljik (1997, 2002) and Bobaljik and Thráinsson (1998) reject the direct causal relation between agreement morphology and verb movement. Instead, they argue that the movement is induced by feature checking for the split-Infl head, Agr and T, and the rich agreement morphology is a result of morphological realisation of those multiple functional heads. I do not evaluate those two competing approaches in this paper due to space limitations.

C in c-structure, as schematised in (11). In the Swedish main clause V2 structure in (12), for instance, the finite verb is base-generated in C and the IP only consists of Spec, which is filled by the subject NP, *han* (cf. Toivonen, 2003, 12).

(11)



(12)



Despite the widespread use of functional categories in the LFG literature, there is not much discussion about the lexical properties of those categories. This paper, by focusing on the variation in verb placement in different languages, aims to uncover the correlation between morphological properties of finite verbs and their categorial status in syntax.

3 Defining richness

The verb agreement features are lexically specified in a verb's lexical entry in LFG. The third singular verb in the present tense in English, for example, can be encoded

as in (13), which states that the verb requires its subject to have attribute-value pairs, ⟨PERS, 3⟩ and ⟨NUM, SG⟩ in f-structure.[4]

(13) *writes* V (↑ PRED) = 'write⟨SUBJ,OBJ⟩'
(↑ TENSE) = PRES
(↑ SUBJ PERS) =$_c$ 3
(↑ SUBJ NUM) =$_c$ SG

In (13), the person feature is specified by using an atomic value, 3. To formulate more fine-grained feature encoding, I adopt a Boolean-valued feature system in this paper (cf. Dalrymple et al., 2009; Sadler, 2011).[5] Under that system, the person features are defined as follows.[6]

(14) 1st: (↑ PERS 1) =$_c$ +
(↑ PERS 2) =$_c$ −

2nd: (↑ PERS 1) =$_c$ −
(↑ PERS 2) =$_c$ +

3rd: (↑ PERS 1) =$_c$ −
(↑ PERS 2) =$_c$ −

Further, I assume that lexical items are paradigmatically organised, so that inflectional forms of the same lexeme compete with each other and the most narrowly specified entry participates in syntax (Paninian Principle or Elsewhere condition) (Anderson 1969; see also Andrews 1982, 1990; Sadler and Spencer 2001; Otoguro 2006, 2014). The upshot of this assumption is that lexical items are not necessarily fully specified for inflectional features, namely syncretic forms have less specific person features than unique forms (see below).

By adopting the Boolean-valued person features and paradigmatic organisation of inflected forms, the lexical entries for Icelandic and Yiddish verbs in Table 1 are formulated as in (15) and (16) respectively.[7]

(15) Icelandic *segja* 'say'

[4]The agreement features can be specified either by costraining equations (=$_c$) or defininining equations (=).

[5]For different feature encoding, **?** use a set value for PERS that consists of S(peaker) and H(earer) as its members.

[6]If a language distinguishes the exclusive first person and the inclusive first person, the inclusive first person is encoded by assigning the positive value for both (↑ PERS 1) and (↑ PERS 2). I assume that languages lacking that distinction have the negative value for the second person feature.

[7]For ease of exposition, PRED is omitted in the entries throughout this paper.

a. *segi* (↑ SUBJ PERS 1) $=_c$ +
(↑ SUBJ PERS 2) $=_c$ −
(↑ SUBJ NUM) $=_c$ SG
(↑ TENSE) = PRES

b. *segir* (↑ SUBJ PERS 1) $=_c$ −
(↑ SUBJ NUM) $=_c$ SG
(↑ TENSE) = PRES

c. *segjum* (↑ SUBJ PERS 1) $=_c$ +
(↑ SUBJ PERS 2) $=_c$ −
(↑ SUBJ NUM) $=_c$ PL
(↑ TENSE) = PRES

d. *segið* (↑ SUBJ PERS 1) $=_c$ −
(↑ SUBJ PERS 2) $=_c$ +
(↑ SUBJ NUM) $=_c$ PL
(↑ TENSE) = PRES

e. *segia* (↑ SUBJ PERS 1) $=_c$ −
(↑ SUBJ PERS 2) $=_c$ −
(↑ SUBJ NUM) $=_c$ PL
(↑ TENSE) = PRES

(16) Yiddish *loyfn* 'run'

a. *loyf* (↑ SUBJ PERS 1) $=_c$ +
(↑ SUBJ PERS 2) $=_c$ −
(↑ SUBJ NUM) $=_c$ SG
(↑ TENSE) = PRES

b. *loyfst* (↑ SUBJ PERS 1) $=_c$ −
(↑ SUBJ PERS 2) $=_c$ +
(↑ SUBJ NUM) $=_c$ SG
(↑ TENSE) = PRES

c. *loyft* (↑ SUBJ PERS 1) $=_c$ −
(↑ TENSE) = PRES

d. *loyfn* (↑ SUBJ PERS 2) $=_c$ −
(↑ SUBJ NUM) $=_c$ PL
(↑ TENSE) = PRES

Since the second singular form and third singular form are syncretic in Icelandic, *sagir* only specifies the value of the first person as −, leaving the second person feature unspecified. All the other forms are fully specific for first and second person features. Following Rohrbacher's insight stated in (10), I assume that the paradigmatic contrast in person in a language's verb inflectional paradigm, i.e. $\langle 1, + \rangle$, $\langle 2, - \rangle$ and $\langle 1, - \rangle$, $\langle 2, + \rangle$, in at least one number of one tense is the condition for finite verbs to be qualified as I. Since the Icelandic verb paradigm (15) satisfies this condition, namely the first person and the second person forms are distinct in the plural in the present tense ((15c) and (15d)), finite verbs are all categorised as I. Similarly, the Yiddish verb paradigm (16) exhibits a full person contrast in the singular of the present tense as shown in (16a) and (16b), so their category is also specified as I.

If we apply the same feature system to finite verbs in Danish in Table 1, however, it has only one entry lacking any specification for person as in (17). As a result, they are categorised as V, not I.

(17) *hørte* (↑ TENSE) = PRES

This approach uniformly captures whether a finite verb in a given language appears in V or I in syntax. Crucially, it is determined lexically and paradigmatically in the language's lexicon.

4 Dialectal variation

It has been reported that the dialectal variation of person markings is found in Hallingmålet in Norway and Älvdalsmålet in Sweden as shown in Table 2.

	Hallingmålet (Norway) *høyræ* 'hear'		Älvdalsmålet (Sweden) *höra* 'hear'	
	Sg	Pl	Sg	Pl
1	høyr-e	høyr-æ	hör-er	hör-um
2	høyr-e	høyr-æ	hör-er	hör-ir
3	høyr-e	høyr-æ	hör-er	hör-a

Table 2: Hallingmålet and Älvdalsmålet verb paradigms (Vikner, 1997, 193)

Although the singular forms are syncretic across different person features in both dialects, Älvdalsmålet has distinct forms for all three person features in the plural. The lexical entries for the verbs in those two dialects are given in (18) and (19). Only the latter exhibits a full person contrast, i.e. (19b) and (19c), which predicts that finite verbs in Älvdalsmålet are I, not V. The relative order between a finite verb and a negation in (21) supports this prediction. Note that (20) confirms that, unlike the Älvdalsmålet dialect, finite verbs in Hallingmålet are V, not I, as predicted from (18).

(18) Hallingmålet *høyræ* 'hear'

a. *høyre* (↑ SUBJ NUM) $=_c$ SG
(↑ TENSE) = PRES

b. *høyræ* (↑ SUBJ NUM) $=_c$ PL
(↑ TENSE) = PRES

(19) Älvdalsmålet *höra* 'hear'

a. *hörer* (↑ SUBJ NUM) $=_c$ SG
(↑ TENSE) = PRES

b. *hörum* (↑ SUBJ PERS 1) $=_c +$
(↑ SUBJ PERS 2) $=_c -$
(↑ SUBJ NUM) $=_c$ PL
(↑ TENSE) = PRES

c. *hörir* (↑ SUBJ PERS 1) $=_c -$
(↑ SUBJ PERS 2) $=_c +$
(↑ SUBJ NUM) $=_c$ PL
(↑ TENSE) = PRES

d. *höra* (↑ SUBJ PERS 1) $=_c -$
(↑ SUBJ PERS 2) $=_c -$
(↑ SUBJ NUM) $=_c$ PL
(↑ TENSE) = PRES

(20) a. *... at me kjøpæ ikkje bokje.
that we buy not the.book
'... that we don't buy the book.'

b. ... at me ikkje kjøpæ bokje.
that we not buy the.book (Hallingmålet)

'. . . that we don't buy the book.'

(21) a. Ba fo ðyæ at ig wild int fy om.
but because that I wanted not follow him
'Just because we didn't want to follow him.'

b. Ig i red an kumb inte.
I am afraid he comes not (Älvdalsmålet)

(Rohrbacher 1999, 118; Bobaljik 2002, 136–7)

Another intriguing variation is found between European Portuguese (EP) and Colloquial Brazilian Portuguese (BP). As Table 3 shows, BP exhibits poorer agreement morphology than EP. In Colloquial BP, the endings of second person singular and first person plural are lost, which makes them syncretic to the second person singular form. Further, while the second person plural and the third person plural are distinct in EP, the former becomes syncretic with the latter in Colloquial BP.

	EP		Colloquial BP	
	Sg	Pl	Sg	Pl
1	falo	falamos	falo	fala
2	falas	falais	fala	falam
3	fala	falam	fala	falam

Table 3: Paradigms of European and Colloquial Brazilian Portuguese verb *falar* 'speak'(Roberts, 2007, 338)

The changes this simplification of inflectional patterns causes are large enough to create a category difference of finite verbs between those two dialects of Portuguese. The lexical entries in (22) and (23) illustrate that EP has unique entries for all person and number combinations, which means that it has a full person feature contrast, whereas Colloquial BP lacks an item with $\langle 1, -\rangle$, $\langle 2, +\rangle$ specification.

(22) EP *falar* 'speak'

a. *falo* (↑ SUBJ PERS 1) = +
(↑ SUBJ PERS 2) = −
(↑ SUBJ NUM) = SG
(↑ TENSE) = PRES

b. *falas* (↑ SUBJ PERS 1) = −
(↑ SUBJ PERS 2) = +
(↑ SUBJ NUM) = SG
(↑ TENSE) = PRES

c. *fala* (↑ SUBJ PERS 1) = −
(↑ SUBJ PERS 2) = −
(↑ SUBJ NUM) = SG
(↑ TENSE) = PRES

d. *falamos* (↑ SUBJ PERS 1) = +
(↑ SUBJ PERS 2) = −
(↑ SUBJ NUM) = PL
(↑ TENSE) = PRES

e. *falais* (↑ SUBJ PERS 1) = −
(↑ SUBJ PERS 2) = +
(↑ SUBJ NUM) = PL
(↑ TENSE) = PRES

f. *falam* (↑ SUBJ PERS 1) = −
(↑ SUBJ PERS 2) = −
(↑ SUBJ NUM) = PL
(↑ TENSE) = PRES

(23) BP *falar* 'speak'

a. *falo* (↑ SUBJ PERS 1) = +
(↑ SUBJ PERS 2) = −
(↑ SUBJ NUM) = SG
(↑ TENSE) = PRES

b. *fala* (↑ TENSE) = PRES

c. *falam* (↑ SUBJ PERS 1) = −
(↑ SUBJ NUM) = PL
(↑ TENSE) = PRES

Those entries suggest that finite verbs in EP are I while those in Colloquial BP are V, and careful observation of the data reveals that the lexical verbs in Colloquial BP stay in V whereas those in EP are located in a higher functional head (Rohrbacher, 1999; Luís and Otoguro, 2012).

Firstly, the finite verb position in EP can be identified by positions of pronominal clitics. EP proclitics attach to a finite verb regardless of whether it is an auxiliary verb or a lexical verb as shown in (24). If the placement of proclitics is consistent in those examples, the auxiliary, *vais*, in (24a) and the lexical verb, *disseram*, in (24b) are both located in I and host the preceding proclitics, *nos* and *me* respectively.

(24) a. Todas as crianças nos disseram a verdade.
all.PL.F the children 1.PL.DAT said the truth
'All the children told us the truth.' (EP)

b. Tu não me vais esquecer.
2.SG not 1.SG.ACC goes forget
'You will not forget me.' (EP)

BP proclitics, on the other hand, only attach to a lexical (thematic) verb as shown in (25), so the proclitic appears between the finite auxiliary and the non-finite lexical verb. Under the assumption that the finite auxiliary, *vai*, is in I and the non-finite lexical verb, *esquecer*, is in V, (25) suggests that proclitics in BP are adjoined to V, not I.

(25) Você vai me esquecer.
2.SG goes 1.SG.ACC forget
'You will forget me.' (BP)

Secondly, to identify the phrase structure position of finite lexical verbs in BP, the location of certain types of adverbs can be used. EP allows a small set of adverbial words such as *ja* 'already', *ainda* 'still' and *não* 'not' to intervene between the clitic and the host as in (26). In (26a), *já* appears between the clitic, *a*, and the host verb, *tens*. (26b) involves the intervention of two adverbs, *ainda* and *não*, between the clitic, *lho*, and the host, *disse*.

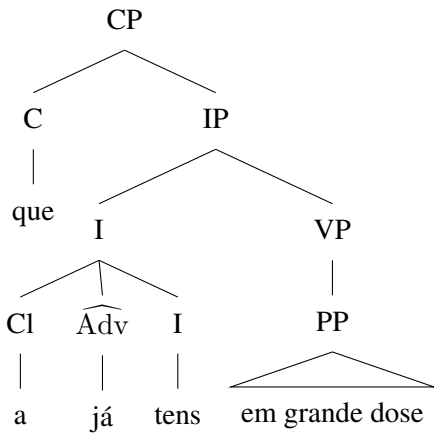
(26) a. ...embora eu saiba que a já tens em grande
...although I know that 3.SG.F.ACC already have in big
dose.
position

'. . . although I know that you already have tons of it (=patience).' (EP)

b. . . . acho que ela lho ainda não disse.
. . . think that she 3.PL.DAT-3.SG.M.ACC yet not told
'. . . I think that s/he hasn't told it to him/her/them yet.' (EP)

Luís and Otoguro (2011) argue that those adverbs are non-projecting words, $\widehat{\text{Adv}}$, and are adjoined to I in EP, so that both proclitics and non-projecting adverbs are adjoined to I as shown in (27).

(27)



The position of *ainda* in the following BP examples (28a) and (28b) suggests that adverbs like *já* and *ainda* are also adjoined to I in BP. The second clause in (28a) consists of a finite auxiliary, *vai*, located in I and a non-finite lexical verb, *comer*, which can be omitted, and the adverb, *ainda*, precedes the finite auxiliary. Similarly, in (28), *ainda* appears immediately before the finite auxiliary verb, *está*. Crucially, unlike EP, those adverbs cannot intervene between proclitic and the host verb as illustrated in (29). In both examples in (29), *já* is adjoined outside of the clitic and the verb complex. Since the proclitic in BP is assumed to be adjoined to V as discussed above, the correct linear ordering obtains if finite lexical verbs in BP are placed in V, not I. (30) represents the c-structure for (29a).

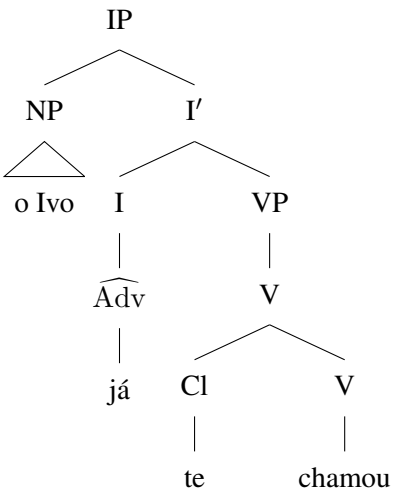
(28) a. Eu já comi, mas a Maria ainda vai [comer].
I already ate-1.SG but the M. still goes eat.INF
'I've already eaten, but Maria's still going to eat.' (BP)
(Nunes and Zocca, 2009, 221)

b. A Bia ainda está fazendo o dever de casa.
the B. still is doing the homework
'Bia is still doing the homework.' (BP) (Silva, 2001, 71)

(29) a. O Ivo já te chamou.
the I. already 2.SG.ACC called
'Ivo has already called you.' (BP)

b. Você já me perguntou?
2.SG already 1.SG.ACC asked
'Have you already asked me?' (BP) (Luís and Otoguro, 2012)

(30)



5 Diachronic change

The present proposal also captures the correlation between diachronic change from a rich morphology to a poor morphology and its effect on verb placement. For instance, the verb inflection in Swedish is very defective like other mainland Scandinavian languages, and finite verbs are placed in V in embedded clauses. However, Old Swedish exhibits a richer inflectional pattern in plural as shown in Table 4. The verb's lexical entries for the inflected forms in Modern Swedish and Old Swedish are given in (31) and (32) respectively. The entries for Old Swedish exhibit a full person contrast, i.e. (32b) and (32c), and predict that the language places a finite verb in I, and this prediction is borne out as shown in (33) where *sivngær* precedes the negation, *ægh*.

	Modern Swedish *bita* 'bite'		Old Swedish *älska* 'love'	
	Sg	Pl	Sg	Pl
1	biter	biter	älsk-ar	älsk-um
2	biter	biter	älsk-ar	älsk-in
3	biter	biter	älsk-ar	älsk-a

Table 4: Modern and Old Swedish verb paradigms (Koeneman and Zeijlstra, 2010, 2012)

(31) Modern Swedish *bita* 'bite'
biter (↑ TENSE) = PRES

(32) Old Swedish *älska* 'love'

a. *älskar* (↑ SUBJ NUM) $=_c$ SG
(↑ TENSE) = PRES

b. *älskum* (↑ SUBJ PERS 1) $=_c$ +
(↑ SUBJ PERS 2) $=_c$ −
(↑ SUBJ NUM) $=_c$ PL
(↑ TENSE) = PRES

c. *älskin* (↑ SUBJ PERS 1) $=_c$ −
(↑ SUBJ PERS 2) $=_c$ +
(↑ SUBJ NUM) $=_c$ PL
(↑ TENSE) = PRES

d. *älska* (↑ SUBJ PERS 1) $=_c$ −
(↑ SUBJ PERS 2) $=_c$ −
(↑ SUBJ NUM) $=_c$ PL
(↑ TENSE) = PRES

(33) . . . æn han sivngær ægh thigianda messu.
if he sings not silent mass
'. . . if he doesn't sing silent mass.'

(Rohrbacher, 1999, 171)

Interestingly, it has been noted that the variety of Swedish spoken in Kronoby, Finland and the Tromsø dialect of Norwegian place finite verbs in I as shown in the ordering between the finite verbs and the negations in (34) and (35), despite their loss of rich inflectional morphology—their verb inflection is as defective as Standard Swedish and Norwegian. I argue that those examples support the lexical approach to the correlation between the verbs' inflectional morphology and their phrase structure positions. As verb's inflectional morphology becomes defective and the person feature contrast has been lost across the various verbs' paradigms, finite verbs become no longer qualified as I. And gradually, the finite verb's category has shifted from I to V. Hence, Kronoby Swedish and Tromsø Norwegian are in the transition stage where the loss of a person feature contrast in verbs' inflectional morphology has been complete, but it has not changed the entire verb category from I to V in the lexicon yet.

(34) He va bra et an tsöfft int bootsen.
it was good that he bought not the.book
'It was good that he didn't buy the book.' (Kronoby Swedish)

(35) . . . før det at han Nilsen kom ikkje.
because that he N. came not
'. . . because Nilsen didn't come.' (Tromsø Norwegian)

(Rohrbacher 1999:118; Bobaljik 2002:139)

In this respect, Faroese shows another interesting property. Its verb inflectional morphology is not fully syncretic, but it retains a unique form for first person singular as shown in Table 5. The lexical entries for those forms are given in (36). Since they do not show a full person feature contrast, we would expect that Faroese verbs

are located in V, not I. However, as (37) illustrates, the language exhibits an intra- or inter-speaker variation with regard to the verb placement: that is a finite verb, *hevði*, can either precede or follow a negation, *ikki*, which suggests that the finite verb can appear either in V or I (see Bobaljik (2002, 140-1) for details). Since Faroese inflectional morphology is not as poor as the mainland Scandinavian languages such as Swedish, Danish and Norwegian, we can assume that it has not completed a category shift from V to I in the lexicon. Therefore, the variation in verb placement among speakers can be observed.

	Faroese *kasta* 'throw'	
	Sg	Pl
1	kasti	kasta
2	kastar	kasta
3	kastar	kasta

Table 5: Faroese verb paradigm (Bobaljik, 2002, 141)

(36) Faroese *kasta* 'throw'

a. *kasti* (↑ SUBJ PERS 1) $=_c$ +
(↑ SUBJ PERS 2) $=_c$ −
(↑ SUBJ NUM) $=_c$ SG
(↑ TENSE) = PRES

b. *kastar* (↑ SUBJ PERS 1) $=_c$ −
(↑ SUBJ NUM) $=_c$ SG
(↑ TENSE) = PRES

c. *kasta* (↑ SUBJ NUM) $=_c$ PL
(↑ TENSE) = PRES

(37) a. Tey nýttu fleiri orð, sum hon hevði ikki hoyrt fyrr.
they used several words which he had not heard before

b. Tey nýttu fleiri orð, sum hon ikki hevði hoyrt fyrr.
they used several words which he not had heard before
'They used several words which he had not heard before.'

(Bobaljik, 2002, 140)

A similar situation is found in Early Modern English (ENE). In Middle English (ME), a finite verb is thought to be located in I and precedes a negation as shown in (38). According to Kroch (1989) and Roberts (1993), the change of verb placement took place in the later 16th and the 17th centuries as illustrated in (39). The loss of rich agreement morphology in ENE, however, took place much earlier, around 1500, as shown in Table 6. Due to the loss of first person singular ending in ENE, the paradigm no longer exhibits a full person feature contrast, namely the first person singular form became identical to the infinitive form, *cast*. Thus, it took a certain amount of time after the loss of rich inflectional morphology to complete

the transition of the entire verb category from V to I in the lexicon.[8]

(38) a. By thy thanks I set not a straw. (Koeneman and Zeijlstra, 2010)
b. if I gave not this accompt to you. (Roberts, 2007, 57)

(39) a. ...he that filches from me my good name robs me of that which not enriches him.
b. Safe on this ground we not fear today to tempt your laughter by our rustic play. (Kroch, 1989, 235)

	ME *cast*		ENE *cast*	
	Sg	Pl	Sg	Pl
1	caste	casten	cast	caste
2	castest	casten	castest	caste
3	casteth	casten	casteth	caste

Table 6: Verb paradigms in Middle English and Early Modern English (cf. Roberts, 1993, 257)

6 Conclusion

The present approach shows that lexical and morphological properties of finite verbs are determining factors for their phrase structure position in syntax. More specifically, the morphological distinction between the first person and the second person features is crucial as discussed in the literature. A Boolean person feature system and paradigmatic organisation of lexical items in the lexicon proposed in this paper allow us to captures such a correlation between the person feature contrast and the verb's categorial properties, i.e. contrast in person features, licenses the verb's categorial status as a functional head. This lexical approach to the interplay between inflectional morphology and verb placement receives further support from the data found in dialectal variation and diachronic change. The analysis illustrates that verb placement can be more naturally accounted for by postulating gradual change of properties in the lexical items, rather than the presence or absence of derivational operations in syntax.

The current proposal provides a framework that can be extended to different types of verb placement in functional categories. As mentioned earlier, Germanic V2 structure involves positing of a finite verb in C. Since CP is a higher projection, which often encodes a clausal level features like mood and modality, different feature encoding must be taken into consideration.

[8] Vikner (1997) shows that a similar time gap between the morphological change and the verb placement change is observed in Middle Danish.

References

Anderson, Stephen R. 1969. *West Scandinavian Vowel Systems and the Ordering of Rules*. Ph. D.thesis, Massachusetts Institute of Technology, Cambridge, MA.

Andrews, Avery D. 1982. The representation of Case in Modern Icelandic. In Joan Bresnan (ed.), *The Mental Representation of Grammatical Relations*, pages 427–503, Cambridge, MA: MIT Press.

Andrews, Avery D. 1990. Unification and Morphological Blocking. *Natural Language and Linguistic Theory* 8, 507–557.

Asudeh, Ash. 2012. *The Logic of Pronominal Resumption*. Oxford: Oxford University Press.

Bobaljik, Jonathan David. 1997. If the Head Fits . . . : On the Morphological Determination of Germanic Syntax. *Linguistics* 35, 1029–1055.

Bobaljik, Jonathan David. 2002. Realizing Germanic Inflection: Why Morphology Does Not Drive Syntax. *Journal of Comparative Germanic Linguistics* 6, 129–167.

Bobaljik, Jonathan David and Thráinsson, Höskuldur. 1998. Two Heads Aren't Always Better Than One. *Syntax* 1, 37–71.

Bresnan, Joan, Asudeh, Ash, Toivonen, Ida and Wechsler, Stephen. 2016. *Lexical-Functional Syntax*. Oxford: Blackwell, second edition.

Dalrymple, Mary, King, Tracy Holloway and Sadler, Louisa. 2009. Indeterminacy by Underspecification. *Journal of Linguistics* 45, 31–68.

Emonds, Joseph E. 1978. The Verbal Complex of V′–V in French. *Linguistic Inquiry* 9, 151–175.

Holmberg, Anders and Platzack, Christer. 1995. *The Role of Inflection in SScandinavian Syntax*. Oxford: Oxford University Press.

King, Tracy Holloway. 1995. *Configuring Topic and Focus in Russian*. Stanford, CA: CSLI Publications.

Koeneman, Olaf and Zeijlstra, Hedde. 2010. Resurrecting the Rich Agreement Hypothesis: Weak isn't Strong Enough. In Dunk-Ho An and Soo-Yeon Kim (eds.), *Movement in Minimalism: Proceedings of the 12th Seoul Conference on Generative Grammar*.

Koeneman, Olaf and Zeijlstra, Hedde. 2012. One Law for the Rich and Another for the Poor: The Rich Agreement Hhypothesis Rehabilitated. lingbuzz/001462.

Kosmeijer, Wim. 1986. The Status of the Finite Inflection in Icelandic and Swe dish. *Working Papers in Scandinavian Syntax* 26, 1–41.

Kroch, Anthony. 1989. Reflexes of Grammar in Patterns of Language Change. *Language Variation and Change* 1, 199–244.

Kroeger, Paul. 1993. *Phrase Structure and Grammatical Relations in Tagalog*. Stanford, CA: CSLI Publications.

Luís, Ana R. and Otoguro, Ryo. 2004. Proclitic contexts in European Portuguese and their effect on clitic placement. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG04 Conference*, pages 334–353, Stanford, CA: CSLI Publications.

Luís, Ana R. and Otoguro, Ryo. 2005. Morphological and Syntactic Well-formedness: The Case of European Portuguese Proclitics. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG05 Conference*, Stanford, CA: CSLI Publications.

Luís, Ana R. and Otoguro, Ryo. 2011. Inflectional Morphology and Syntax in Correspondence: Evidence from European Portuguese. In Alexandra Galani, Glyn Hicks and George Tsoulas (eds.), *Morphology and Its Interfaces*, pages 187–225, Amsterdam: John Benjamins.

Luís, Ana R. and Otoguro, Ryo. 2012. Non-Isomorphism and Microvariation: A Comparative Analysis of Pronominal Clitics in European and Brazilian Portuguese. Paper presented at Clitics and Beyond, University of Göttingen, Germany.

Nunes, Jairo and Zocca, Cynthia. 2009. Lack of Morphological Identity and Ellipsis Resolution in Brazilian Portuguese. In Jairo Nunes (ed.), *Minimalist Essays on Brazilian Portuguese*, pages 215–236, Amsterdam: John Benjamins.

Otoguro, Ryo. 2006. *Morphosyntax of Case: A Theoretical Investigation of the Concept*. Ph. D.thesis, University of Essex.

Otoguro, Ryo. 2014. Constructional Paradigm in Constraint-based Morphosyntax: A Case of Japanese Verb Inflection. In Kayla Carpenter, Oana David, Florian Lionnet, Christine Sheil, Tammy Stark and Vivian Wauters (eds.), *Proceedings of the Annual Meeting of the Berkeley Linguistic Society 38*, pages 371–386, Berkeley Linguistic Society, Berkeley, CA.

Pollock, Yves. 1989. Verb Movement, Universal Grammar, and the Structure of IP. *Linguistic Inquiry* 20, 365–424.

Roberts, Ian. 1993. *Verbs and Diachronic Syntax*. Dordrecht: Kluwer Academic Publisher.

Roberts, Ian. 2007. *Diachronic Syntax*. Oxford: Oxford University Press.

Rohrbacher, Bernhard Wolfgang. 1999. *Morphology-driven Syntax: A Theory of V to I Raising and Pro-drop*. Amsterdam: John Benjamins.

Sadler, Louisa. 1997. Clitics and the structure-function mapping. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG97 Conference*, Stanford, CA: CSLI Publications.

Sadler, Louisa. 2011. Indeterminacy, Complex Features and Underspecification. *Morphology* 21, 379–417.

Sadler, Louisa and Spencer, Andrew. 2001. Syntax as an exponent of morphological features. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 2000*, pages 71–96, Dordrecht: Kluwer Academic Publishers.

Sells, Peter. 2001. *Structure, Alignment and Optimality in Swedish*. Stanford, CA: CSLI Publications.

Sells, Peter. 2003. Stylistic Fronting in Icelandic: A Base-generated Construction. *Gengo Kenkyu* 123, 257–297.

Sells, Peter. 2005. The Peripherality of the Icelandic Expletive. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG05 Conference*, pages 408–428, Stanford, CA: CSLI Publications.

Silva, Gláucia Valeria. 2001. *Word Order in Brazilian Portuguese*. Berlin: Mouton de Gruyter.

Toivonen, Ida. 2003. *Non-Projecting Words: A Case Study of Swedish Particles*. Dordrecht: Kluwer Academic Publishers.

Travis, Lisa. 1991. *Parameters and Effects of Word Order Variation*. Ph. D.thesis, Massachusetts Institute of Technology, Cambridge, MA.

Vikner, Sten. 1995. *Verb Movement and Expletive Subjects in the Germanic Languages*. Oxford: Oxford University Press.

Vikner, Sten. 1997. V^0-to-I^0 Movement and Inflection fro Person in All Tenses. In Liliane Haegeman (ed.), *The New Comparative Syntax*, pages 189–213, London: Longman.

AN LFG ANALYSIS OF THE SO-CALLED REFLEXIVE MARKER IN POLISH

Agnieszka Patejuk and Adam Przepiórkowski
Institute of Computer Science, Polish Academy of Sciences

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King (Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

This paper proposes an LFG analysis of SIĘ – a Polish word which is usually referred to as the reflexive pronoun even though it is not always the case that it is used in this function. Moreover, SIĘ can be used in more than one function at the same time; there is also the issue of haplology of SIĘ in verb chains and of sharing it under coordination – the proposed analysis captures these phenomena. The paper discusses the status of SIĘ, whether it is an argument or a marker (a co-head with no PRED), and treats it as a marker, though some counterevidence against a unified analysis is presented.

1 Introduction

This paper provides an LFG analysis of SIĘ, a Polish word which is typically referred to as a "reflexive" item (e.g. Kupść 2000, Przepiórkowski et al. 2002, Kibort 2004), even though this is only one of its possible functions. It aims to provide a comprehensive treatment (within the available space limits) instead of focusing on a selected aspect of this phenomenon. The analysis presented here has been implemented as a part of a large XLE grammar of Polish (Patejuk and Przepiórkowski 2012).

The initial sections of this paper present the linguistic data on the basis of attested examples taken from the National Corpus of Polish (*Narodowy Korpus Języka Polskiego*, NKJP; Przepiórkowski et al. 2012, `http://nkjp.pl/`) or retrieved using Google search engine.[1] § 2 presents selected basic types of SIĘ in Polish and briefly discusses their properties. § 3 shows that SIĘ may bear more than one function in relation to the same predicate. § 4 presents the phenomenon of haplology of SIĘ in verb chains, while § 5 discusses sharing of SIĘ under coordination of verbal predicates.

§ 6 is devoted to the issue of the status of SIĘ: whether it should be represented as an argument of the relevant predicate or as a marker, i.e. a co-head that does not have its own PRED value, but only contributes certain features to the f-structure of the predicate it depends on. It presents the proposed analysis and formalisation of phenomena discussed earlier based on the latter representation of SIĘ – as a marker. However, this issue is discussed further in § 7, which shows potential problems with such an analysis of SIĘ. Finally, § 8 concludes the paper.

It is worth noting that Polish SIĘ has been described earlier in Kupść 2000, which provides a comprehensive discussion of rich data and an HPSG analysis for phenomena such as multifunctionality of SIĘ and haplology of SIĘ in verb chains.[2] However, the current analysis is not an LFG reimplementation of Kupść 2000 – it offers a different solution for the representation of SIĘ (which provides detailed information about the type of SIĘ involved, providing an explicit account of ambiguous readings with SIĘ); it also includes an analysis of sharing SIĘ under coordination, which is not included in Kupść 2000.

[1] All links provided in this paper were accessed on 20 December 2015.

[2] Sharing of SIĘ, though discussed there, was not formalised.

2 Basic uses of SIĘ

Examples below show selected basic uses of SIĘ: reflexive (REFL) in (1), reciprocal (RECIP) in (2), inherent (INH) in (3), impersonal (IMPRS) in (4).[3,4]

(1) Jacek golił się.
Jacek.SG.M1 shaved.SG.M1 REFL
'Jacek was shaving.' (NKJP)

(2) Przez chwilę całowali się.
for moment kissed.PL.M1 RECIP
'For a moment, they were kissing each other.' (NKJP)

(3) Ty też się śmiejesz.
you.SG also INH laugh.SG.2
'You're also laughing.' (NKJP)

(4) Po Edenie chodziło się nago.
along Eden walked.3.SG.N IMPRS nude.ADV
'One would walk nude in Eden.' (NKJP)

Other uses of SIĘ, not discussed in this paper, include middle (which was, however, discussed in Kupść 2000), shown below (glossed as MID):

(5) Takie powieści zawsze dobrze się sprzedawały.
such.PL.F novel.PL.F always well MID sold.PL.F
'Such novels have always sold well.' (NKJP)

While all types of SIĘ listed in (1)–(4) above have the same shape, the following subsections discuss their properties with respect to criteria such as the possibility of being replaced with SIEBIE (§ 2.1), being required lexically or being constructional (§ 2.2), making a contribution to the semantics (§ 2.3) and finally imposing extra constraints, for instance on verbal agreement (§ 2.4).

2.1 SIĘ vs. SIEBIE

Though SIĘ is typically described as a weak form of the pronoun SIEBIE 'self' (e.g. Kupść 2000, but also in more traditional linguistic literature), it can only be replaced with appropriate forms of SIEBIE in (1)–(2), which feature reflexive and reciprocal SIĘ, respectively:

(6) Jacek golił siebie (samego).
Jacek.SG.M1 shaved.SG.M1 self.ACC own.ACC
'Jacek was shaving himself.'

(7) Przez chwilę całowali siebie (nawzajem).
for moment kissed.PL.M1 self.ACC reciprocally
'For a moment, they were kissing each other.'

[3]Though the word *nago* is translated as 'nude' in (4), it is an adverb – as indicated in glosses.

[4]Morphosyntactic abbreviations mostly adhere to those recommended in Leipzig Glossing Rules (`http://www.eva.mpg.de/lingua/resources/glossing-rules.php`), with the exception of M1, which refers to the human-masculine gender (assuming the repertoire of five genders in Polish proposed in Mańczak 1956).

(8) *Ty też siebie śmiejesz.

(9) *Po Edenie chodziło siebie nago.

Though, as shown above, certain types of SIĘ can be replaced with SIEBIE, it does not follow that these forms are equivalent and that they have the same syntactic properties – this issue is discussed in more detail in § 6.

2.2 Required lexically vs. constructional

Some types of SIĘ are a lexical requirement of the given predicate: these include reflexive (cf. (1)), reciprocal (cf. (2)) and inherent (cf. (3)) SIĘ.

By contrast, impersonal SIĘ shown in (4) is not required lexically – it is a productive construction, which may be used with almost any predicate that allows for a human subject. This constraint is illustrated below: since the subject of BOLEĆ 'hurt' in (10) is not human, the use of impersonal SIĘ is ungrammatical.

(10) Bolało mnie (kolano).
hurt.SG.N I.ACC knee.NOM.SG.N
'My knee would (often) hurt.'

(11) *Bolało mnie się.
hurt.SG.N I.ACC IMPRS
'It would (often) hurt.' (intended)

2.3 Contribution to semantics

Another property to be taken into account is whether the given type of SIĘ contributes to the semantics – it is the case with all types of SIĘ discussed above except for inherent SIĘ, which is required syntactically but it does not make any contribution to the semantics. By contrast, the remaining types of SIĘ make such a contribution: the reflexive SIĘ marks that the respective predicate is reflexive, the reciprocal SIĘ marks that it is reciprocal, influencing the semantic interpretation, while the impersonal SIĘ marks that the predicate is impersonal.

2.4 Agreement constraints

When the impersonal SIĘ is used, it triggers default agreement on the verb (Dziwirek 1990) – it must be third person, singular, neuter, though it must be noted that neuter gender specification is only visible with past tense forms such as in (12), but not with a present tense form such as in (13).

(12) Po Edenie chodziło/*chodziły się nago.
along Eden walked.3.SG/PL.N IMPRS nude.ADV
'One would walk nude in Eden.' (NKJP)

(13) Po Edenie chodzi/*chodzą się nago.
along Eden walk.3.SG/PL IMPRS nude.ADV
'One walks nude in Eden.'

Furthermore, as shown in (14), the impersonal SIĘ blocks the use of a lexical subject. It is the case even when the lexical subject would normally be compatible with such a form (third person, singular, neuter), compare (15).

(14) Po Edenie (*każde dziecko) chodziło się nago.
along Eden every child.NOM.SG.N walked.3.SG.N IMPRS nude.ADV

(15) Po Edenie każde dziecko chodziło nago.
along Eden every child.NOM.SG.N walked.3.SG.N nude.ADV
'Every child walked nude in Eden.'

3 Multifunctional use of SIĘ

It is possible for SIĘ to have more than one function with respect to the same predicate at the same time: in both examples below SIĘ is impersonal (as in (4)), apart from being inherent in (16) (as in (3)) and reflexive in (17) (as in (1)). Multifunctionality of SIĘ is represented in glosses by joining appropriate basic types of SIĘ using the + symbol.

(16) Kiedyś śmiało się z czerwonych.
sometime laughed.SG.N INH+IMPRS from red.PL
'One used to laugh at the communists.' (Google)[5]

(17) Kiedyś goliło się żyletkami.
sometime shaved.SG.N REFL+IMPRS razor blade.PL.INST
'One used to shave with razor blades.' (Google)[6]

Furthermore, as noted in Kupść 2000, § 3.5.1, it is ungrammatical to use more than one instance of SIĘ as dependents of the same predicate, so the use of multifunctional SIĘ is the only way to obtain the readings presented in (16)–(17) above:

(18) *Kiedyś się śmiało się z czerwonych.
sometime IMPRS laughed.SG.N INH from red.PL

(19) *Kiedyś się goliło się żyletkami.
sometime IMPRS shaved.SG.N REFL razor blade.PL.INST

It is perhaps worth mentioning that although the interpretations provided in (16)–(17) are the intended readings, there are alternative readings of these sentences which depend on the interpretation of SIĘ. (16) has a reading where SIĘ is not multifunctional but only inherent – under this reading the subject is implicit and, because of the agreement properties of the verb, whose form is third person singular neuter, it could, for instance, refer to a child, as shown in (20):

(20) Kiedyś (dziecko) śmiało się z czerwonych.
sometime child.NOM.SG.N laughed.SG.N INH from red.PL
'Once a child laughed at the communists.'

[5] http://forum.wirtualnemedia.pl/teologowie-moralisci-apeluja-do-mediow-t9375.html

[6] https://odjechani.com.pl/Thread-Pierwsze-golenie?page=2

(17) in turn has two alternative readings: one where SIĘ is only reflexive and the subject is implicit (as discussed for (20)), shown in (21), and the other where SIĘ is only impersonal (with an implicit subject that blocks the use of a lexical one, see § 2.4) and the predicate GOLIĆ takes an implicit object, see (22):

(21) Kiedyś (dziecko) goliło się żyletkami.
sometime child.NOM.SG.N shaved.SG.N REFL razor blade.PL.INST
'A child used to shave with razor blades.'

(22) Kiedyś goliło się (ludzi) żyletkami.
sometime shaved.SG.N IMPRS people.ACC razor blade.PL.INST
'One used to shave people with razor blades.'

All readings presented here are accounted for by the proposed analysis.

4 Haplology of SIĘ in verb chains

When two (or more) predicates in a verb chain consisting of the main verb and any nonempty sequence of subordinate infinitival clauses require SIĘ, the one belonging to the structurally higher predicate can at the same time satisfy the requirements of the lower one – as a result, it is possible to use only one instance of SIĘ, as in (23).[7] Such sharing of SIĘ in verb chains is represented in glosses by joining the types of SIĘ required by respective verbs using the / symbol.

(23) A czy Tobie zdarzyło się śmiać z dowcipu który
and Q you.DAT happened.SG.N INH/INH laugh.INF from joke which
nie był [...] śmieszny?
NEG was funny
'Have you happened to laugh at a joke that was not funny?' (Google)[8]

In (23) both predicates – ZDARZYĆ 'happen' and ŚMIAĆ 'laugh' – require the inherent SIĘ (as shown in (24)), but there is only one textual instance of SIĘ, which structurally belongs to the higher predicate, namely ZDARZYĆ, as demonstrated in (25), which is ungrammatical with SIĘ placed in the lower clause.

(24) A czy Tobie zdarzyło się śmiać się z dowcipu który
and Q you.DAT happened.SG.N INH laugh.INF INH from joke which
nie był [...] śmieszny?
NEG was funny

(25) *A czy Tobie zdarzyło śmiać się z dowcipu który
and Q you.DAT happened.SG.N laugh.INF INH/INH from joke which
nie był [...] śmieszny?
NEG was funny

The following examples show that haplology of SIĘ is also possible when, unlike in (23), the respective verbs in the verb chain require different types of SIĘ:

[7] CZY, glossed as Q in (23)–(25), is the Polish *yes/no* question particle.

[8] http://www.druga-strona.pl/oobe-podroz-poza-cialem/czy-warto-zdobyc-kazda-wiedze-na-temat-oobe

(26) Sporo osób boi się golić pod włos.
many people fear.SG.N INH/REFL shave.INF under hair
'Many people are afraid of shaving against the grain.' (Google)[9]

(27) Sporo osób boi się golić się pod włos.
many people fear.SG.N INH shave.INF REFL under hair

(28) Kiedyś chodziło się do kina gapić na wielki mrugający
sometime walked.SG.N IMPRS/INH to cinema stare.INF at great blinking
ekran.
screen
'One would go to a cinema to stare at the great blinking screen.' (NKJP)

(29) Kiedyś chodziło się do kina gapić się na wielki mrugający
sometime walked.SG.N IMPRS to cinema stare.INF INH at great blinking
ekran.
screen

In (26) the main verb, BAĆ 'fear', requires the inherent SIĘ, while its infinitival complement, namely GOLIĆ 'shave', takes the reflexive SIĘ – as shown in (27). Similarly, in (28), the main verb CHODZIĆ 'walk' forms a construction with the impersonal SIĘ, while the embedded predicate GAPIĆ 'stare' requires the inherent SIĘ – as demonstrated in (29).

Finally, as mentioned when discussing multifunctionality of SIĘ, sentences with SIĘ may be ambiguous depending on the interpretation of SIĘ. This is the case in (26), which has has an alternative reading presented in (30) where haplology is not involved – SIĘ is only a dependent of the main verb (which takes inherent SIĘ) and SHAVE has an implicit object (as in (22)):

(30) Sporo osób boi się golić (ludzi) pod włos.
many people fear.SG.N INH shave.INF people.ACC under hair
'Many people are afraid to shave people against the grain.'

This reading is also taken into account under the proposed analysis.

5 Sharing SIĘ under coordination

Like other dependents, SIĘ can be shared by coordinated predicates: not only in the typical situation when it is one of the edge constituents, as in (31), but also when it is "intertwined" inside one of the conjuncts, as shown in (32), where it would normally belong to the phrase headed by CAŁOWAĆ. Such sharing is represented in glosses by joining types of SIĘ required by the coordinated predicates using |.

(31) Całowali i przytulali się w metrze.
kissed.PL.M1 and hugged.PL.M1 RECIP|RECIP in subway
'They were kissing and hugging each other in subway.' (Google)[10]

[9] http://www.brzytwa.org/forum/index.php?topic=2660.0
[10] http://warszawa.wyborcza.pl/warszawa/51,34889,9731392.html

(32) Całowali się i przytulali!
kissed.PL.M1 RECIP|RECIP and hugged.PL.M1
'They were kissing and hugging each other!' (Google)[11]

(33) Całowali się i przytulali się!
kissed.PL.M1 RECIP and hugged.PL.M1 RECIP

As it is the case in verb chains, while SIĘ may have the same function with both coordinated verbs – see (31)–(32), where CAŁOWAĆ 'kiss (each other)' and PRZYTULAĆ 'hug (each other)' both require the reciprocal SIĘ (as shown in (33)) – it may also be different for each conjunct – compare (34), where SIĘ required by ŚMIAĆ 'laugh' is inherent, while with the figurative PUKAĆ SIĘ W GŁOWĘ, lit. 'knock yourself in head', i.e., 'imply that somebody is nuts', it is reflexive (see (35)).

(34) Śmiali się i pukali w głowy.
laughed.PL.M1 INH|REFL and knocked.PL.M1 in heads
'They were laughing and asking if somebody is nuts.' (NKJP)

(35) Śmiali się i pukali się w głowy.
laughed.PL.M1 INH and knocked.PL.M1 REFL in heads

6 Representation of SIĘ: a marker

Though, as mentioned in § 2.1, some types of SIĘ can be replaced with SIEBIE, even there the referential properties of the two forms differ in the sense of Jackendoff 1992 – the following examples, whose context is a photograph of Zlatan Ibramović next to the wax figure of himself in the Musée Grévin, show differences in meaning transfer between these two forms:

(36) Zlatan podziwiał się w muzeum.
Zlatan admired REFL in museum
'Zlatan admired himself in the museum'.

(37) Zlatan podziwiał siebie w muzeum.
Zlatan admired self.ACC in museum
'Zlatan admired himself/the wax figure of himself in the museum'.

In (36) the use of SIĘ seems to force strict identity: it can only mean that Zlatan admired himself (rather than the wax figure of himself). By contrast, (37) has two readings: one is the reading with identity by co-reference where Zlatan admires himself (same as in (36)), while the other is the sloppy reading where Zlatan admires the wax figure of himself (unavailable in (36)). This difference suggests that while SIEBIE is an argument and there are 2 syntactic participants in sentences with SIEBIE (subject and object, see (39)), SIĘ is rather a marker, a co-head that does not have a PRED value of its own, so there is only one participant (the subject, as in (38)), which would explain the requirement of the strict identity reading in (36).

[11] http://www.pudelek.pl/artykul/63056/rozenek_i_majdan_razem_w_mcdonaldzie_calowali_sie_i_przytulali/

(38)
$$\begin{bmatrix} \text{PRED} & \text{`ADMIRE}\langle \boxed{1} \rangle\text{'} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix} \text{PRED} & \text{`ZLATAN'} \\ \text{CASE} & \text{NOM} \end{bmatrix} \end{bmatrix}$$

(39)
$$\begin{bmatrix} \text{PRED} & \text{`ADMIRE}\langle \boxed{1}, \boxed{2} \rangle\text{'} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix} \text{PRED} & \text{`ZLATAN'} \\ \text{CASE} & \text{NOM} \end{bmatrix} \\ \text{OBJ} & \boxed{2}\begin{bmatrix} \text{PRED} & \text{`SELF'} \\ \text{CASE} & \text{ACC} \end{bmatrix} \end{bmatrix}$$

The formalisation of the proposed analysis of SIĘ is inspired by the analysis of case offered in Dalrymple et al. 2009, which involves the use of a complex CASE attribute containing subattributes corresponding to particular values of case, each of which takes a boolean value, thereby making it possible to account for multiple case marking of a dependent in languages such as Polish, where different predicates can impose different case requirements on the shared dependent. This idea can be adapted to account for the multifunctionality of SIĘ discussed in § 3.

6.1 Templates used in lexical entries

The template called in the lexical entry of SIĘ is provided in (40):

(40) $\text{SIE} \equiv (\uparrow \text{SIE PRESENT}) = + \ \wedge \ [\text{SIE-LEX-IMPRS} \ \vee \ \text{SIE-IMPRS-ONLY}]$

(41) $\text{SIE-LEX-IMPRS} \equiv \text{SIE-LEX} \wedge (\text{SIE-IMPRS})$

(42) $\text{SIE-IMPRS-ONLY} \equiv \text{SIE-IMPRS} \wedge \neg\text{SIE-LEX}$

(43) $\text{SIE-LEX} \equiv (\uparrow \text{SIE} \{\text{INH}|\text{REFL}|\text{RECIP}\}) =_c +$

(44) $\text{SIE-IMPRS} \equiv (\uparrow \text{SIE IMPRS}) = + \ \wedge \ (\uparrow \text{SUBJ}) = \%\text{S} \ \wedge \ (\%\text{S PRED}) = \text{`PRO'} \ \wedge \ (\%\text{S PERS}) = 3 \ \wedge \ (\%\text{S NUM}) = \text{SG} \ \wedge \ (\%\text{S GEND}) = \text{N}$

The definition of the SIE template consists of two conjoined statements. The first one introduces the attribute SIE, which contains the attribute PRESENT, whose value is set to $+$: it marks the presence of SIĘ, which is analysed as a co-head of the relevant verb, so it contributes to its f-structure. In order to account for the fact that only one instance of SIĘ may be used with the given predicate, as discussed in § 3, the attribute PRESENT is an instantiated feature whose value may be set only once.

The second conjunct of (40) is a disjunction of two template calls; let us start with the discussion of its first disjunct. SIE-LEX-IMPRS, whose definition is given in (41), rewrites to a conjunction of two template calls: the first one, SIE-LEX defined in (43), checks whether the attribute SIE has the positive value for any of the relevant types of SIĘ (reflexive, reciprocal or inherent), which accounts for lexically required SIĘ, while the second one is the optional call (put in brackets) to the template SIE-IMPRS defined in (44). If this call is used, it adds the attribute IMPRS to the value of the SIE attribute, making SIĘ multifunctional: it is both impersonal and of one of the other three types listed above (see § 3). Additionally, SIE-IMPRS imposes constraints on the subject of the verb to which it attaches: it introduces a pro subject to block the occurrence of a lexical one and it imposes appropriate agreement constraints (third person, singular number, neuter gender).

Let us now discuss the template SIE-IMPRS-ONLY called in the second disjunct of the second conjunct of (40), defined in (42) as a conjunction of two template calls: the first conjunct is an obligatory (not in brackets) call to the template SIE-IMPRS defined in (44) and discussed above, which introduces the constructional impersonal SIĘ, while the second conjunct is a negated (¬) call to the SIE-LEX template – its result is precluding SIĘ from being of any of the types defined in (43), which includes inherent, reflective and reciprocal SIĘ. The effect of these constraints is that SIĘ introduced by SIE-IMPRS-ONLY cannot be multifunctional – it can only be impersonal.

The last part of the analysis is placed in the lexical entries of verbs requiring SIĘ – they contain calls to the template defined in (45), where the value of the parameter TYPE corresponds to the type of required SIĘ (INH, REFL or RECIP; impersonal SIĘ of IMPRS type is not included here as it is constructional, as discussed in § 2.2):

(45) $\text{SIE-TYPE(TYPE)} \equiv (\uparrow \text{SIE TYPE}) = + \ \wedge$
$((\text{XCOMP}^* \uparrow) \text{ SIE PRESENT}) =_c +$

The first conjunct of (45) specifies the type of required SIĘ, while the second one ensures that SIĘ is present in the relevant domain – it looks for the positive value of the PRESENT attribute inside the SIE attribute locally to the verb (when XCOMP* in (45) corresponds to zero instances of XCOMP, the path then points to the local verb) or higher in the verb chain (when the local verb is nested in a structure containing one or more successive XCOMPs), which accounts for the haplology of SIĘ discussed in § 4. Sample lexical entries for verbs used in examples (1)–(3) are provided below:

(46) *golić* V $(\uparrow \text{PRED}) = \text{‘SHAVE<}(\uparrow \text{SUBJ})\text{>’} \wedge @(\text{SIE-TYPE REFL})$

(47) *całować* V $(\uparrow \text{PRED}) = \text{‘KISS<}(\uparrow \text{SUBJ})\text{>’} \wedge @(\text{SIE-TYPE RECIP})$

(48) *śmiać* V $(\uparrow \text{PRED}) = \text{‘LAUGH<}(\uparrow \text{SUBJ})\text{>’} \wedge @(\text{SIE-TYPE INH})$

6.2 Simple structures

The following examples provide simplified f-structures for sentences presented in § 2: (49) and (50) correspond to (1) and (2), which feature reflexive (REFL) and reciprocal (RECIP) SIĘ, respectively. Due to the adopted analysis of SIĘ as a marker that does not have its own PRED attribute (see § 6), these f-structures feature one-argument (rather than two-argument) predicates SHAVE and KISS.

(49)
$$\begin{bmatrix} \text{PRED} & \text{‘SHAVE}\langle \boxed{1} \rangle\text{’} \\ \text{SUBJ} & \boxed{1} \begin{bmatrix} \text{PRED} & \text{‘JACEK’} \\ \text{CASE} & \text{NOM} \\ \text{NUM} & \text{SG} \end{bmatrix} \\ \text{SIE} & \begin{bmatrix} \text{REFL} & + \\ \text{PRESENT} & + \end{bmatrix} \end{bmatrix}$$

(50)
$$\begin{bmatrix} \text{PRED} & \text{‘KISS}\langle \boxed{1} \rangle\text{’} \\ \text{SUBJ} & \boxed{1} \begin{bmatrix} \text{PRED} & \text{‘PRO’} \\ \text{CASE} & \text{NOM} \\ \text{NUM} & \text{PL} \end{bmatrix} \\ \text{SIE} & \begin{bmatrix} \text{RECIP} & + \\ \text{PRESENT} & + \end{bmatrix} \end{bmatrix}$$

The representation of the remaining two types of SIĘ is not controversial: (51) corresponds to (3), which features inherent (INH) SIĘ, while (52) provides a sim-

plified representation of (4), which contains an impersonal (IMPRS) SIĘ.

(51)
$$\begin{bmatrix} \text{PRED} & \text{`LAUGH}\langle \boxed{1} \rangle\text{'} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix} \text{PRED} & \text{`PRO'} \\ \text{CASE} & \text{NOM} \\ \text{NUM} & \text{SG} \\ \text{PERS} & 2 \end{bmatrix} \\ \text{SIE} & \begin{bmatrix} \text{INH} & + \\ \text{PRESENT} & + \end{bmatrix} \end{bmatrix}$$

(52)
$$\begin{bmatrix} \text{PRED} & \text{`WALK}\langle \boxed{1} \rangle\text{'} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix} \text{PRED} & \text{`PRO'} \\ \text{GEND} & \text{N} \\ \text{NUM} & \text{SG} \\ \text{PERS} & 3 \end{bmatrix} \\ \text{SIE} & \begin{bmatrix} \text{IMPRS} & + \\ \text{PRESENT} & + \end{bmatrix} \end{bmatrix}$$

6.3 Multifunctional SIĘ

Let us now consider some structures produced by this analysis for sentences that involve multifunctional SIĘ discussed in § 3 – the f-structures in (53)–(54) provide a representation of simplified versions of (16) (*śmiało się z czerwonych*) and (17) (*goliło się żyletkami*).

(53)
$$\begin{bmatrix} \text{PRED} & \text{`LAUGH}\langle \boxed{1}, \boxed{2} \rangle\text{'} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix} \text{PRED} & \text{`PRO'} \end{bmatrix} \\ \text{OBL} & \boxed{2}\begin{bmatrix} \text{PRED} & \text{`RED'} \\ \text{PFORM} & \text{Z} \\ \text{CASE} & \text{GEN} \end{bmatrix} \\ \text{SIE} & \begin{bmatrix} \text{INH} & + \\ \text{IMPRS} & + \\ \text{PRESENT} & + \end{bmatrix} \end{bmatrix}$$

(54)
$$\begin{bmatrix} \text{PRED} & \text{`SHAVE}\langle \boxed{1} \rangle\text{'} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix} \text{PRED} & \text{`PRO'} \end{bmatrix} \\ \text{ADJ} & \left\{ \begin{bmatrix} \text{PRED} & \text{`RAZOR'} \\ \text{CASE} & \text{INST} \end{bmatrix} \right\} \\ \text{SIE} & \begin{bmatrix} \text{REFL} & + \\ \text{IMPRS} & + \\ \text{PRESENT} & + \end{bmatrix} \end{bmatrix}$$

In (16) SIĘ is at the same time inherent (specified lexically by LAUGH) and impersonal (constructionally), while in (17) SIĘ is both reflexive (as specified by SHAVE) and impersonal.

Such multifunctional specifications are the result of the interaction of lexical entries of verbs requiring SIĘ lexically and the lexical entry of SIĘ (see (40)), which contains a call to the template SIE-LEX-IMPRS defined in (41) that can optionally introduce the specification of the impersonal SIĘ (given in (44)) if it satisfies the condition formalised in (43), namely that the given instance of SIĘ is already specified lexically by the appropriate verb (using (45)) as belonging to one of the following types: INH (as in (53)), REFL (as in (54)) or RECIP.

6.4 Haplology of SIĘ

Under the current analysis, SIĘ contributes the attribute PRESENT inside the SIE attribute of the relevant verb (as a result of calling the template defined in (40)), while verbs whose lexical requirement of SIĘ is satisfied as a result of haplology do not have this attribute, which makes it possible to identify instances of SIĘ haplology

in f-structures. Let us discuss the f-structure provided in (55), which corresponds to a simplified version of (26) (*Sporo boi się golić*) and uses this mechanism:

(55)
$$\begin{bmatrix} \text{PRED} & \text{`FEAR}\langle\boxed{1},\boxed{2}\rangle\text{'} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix}\text{PRED} & \text{`MANY'}\end{bmatrix} \\ \text{XCOMP} & \boxed{2}\begin{bmatrix}\text{PRED} & \text{`SHAVE}\langle\boxed{1}\rangle\text{'} \\ \text{SIE} & \begin{bmatrix}\text{REFL} & +\end{bmatrix}\end{bmatrix} \\ \text{SIE} & \begin{bmatrix}\text{INH} & + \\ \text{PRESENT} & +\end{bmatrix} \end{bmatrix}$$

The higher predicate, FEAR, to which SIĘ belongs structurally, contains two attributes inside SIE: PRESENT, contributed by the lexical instance of SIĘ, and INH, which corresponds to the type of SIĘ required lexically by FEAR (with the help of the template defined in (45)). By contrast, the SIE attribute of the infinitival complement (XCOMP) only contains the REFL specification of SIĘ which is required by SHAVE – since this predicate does not have a lexical SIĘ as its dependent but it uses haplology to satisfy this requirement, there is no PRESENT attribute inside its SIE attribute. It is worth noting that, since it is the verb that determines the type of SIĘ (using a defining equation in (45)), it is possible to handle haplology where different types of SIĘ are involved, as in (26) discussed above, where the main verb takes the inherent SIĘ, while its infinitival complement takes the reflexive SIĘ. Similarly, in (28) SIĘ belonging to the main verb is impersonal (constructional, not required lexically as in (26)), while the infinitival complement takes the inherent SIĘ.

Finally, such an analysis of haplology ensures that the requirement of an appropriate type of SIĘ of relevant predicates is satisfied without placing the SIE attribute in the f-structures of all subsequent predicates in the verb chain which may have no such requirement, as in (56), whose f-structure is provided in (57) – as shown in (58), unlike BAĆ 'fear' and ŚMIAĆ 'laugh', PRÓBOWAĆ 'try' does not require SIĘ (this is represented using ... in glosses in (56)) and putting the SIE attribute in the corresponding partial f-structure would distort its semantics.

(56) Bał się próbować uśmiechnąć.
feared.SG.M1 INH/.../INH try.INF smile.INF
'He was afraid to try to smile.'

(57)
$$\begin{bmatrix} \text{PRED} & \text{`FEAR}\langle\boxed{1},\boxed{2}\rangle\text{'} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix}\text{PRED} & \text{`PRO'}\end{bmatrix} \\ \text{XCOMP} & \boxed{2}\begin{bmatrix}\text{PRED} & \text{`TRY}\langle\boxed{3}\rangle\boxed{1}\text{'} \\ \text{SUBJ} & \boxed{1} \\ \text{XCOMP} & \boxed{3}\begin{bmatrix}\text{PRED} & \text{`SMILE}\langle\boxed{1}\rangle\text{'} \\ \text{SIE} & \begin{bmatrix}\text{INH} & +\end{bmatrix}\end{bmatrix}\end{bmatrix} \\ \text{SIE} & \begin{bmatrix}\text{INH} & + \\ \text{PRESENT} & +\end{bmatrix} \end{bmatrix}$$

(58) Bał się próbować (*się) uśmiechnąć się.
feared.SG.M1 INH try.INF INH smile.INF INH

6.5 Sharing SIĘ under coordination

The last remaining issue is sharing of SIĘ under coordination. While examples where SIĘ is one of the edge constituents, as in (31), can be handled using standard mechanisms for dependent sharing under coordination, such a solution is not sufficient to account for more difficult cases such as the ones provided in (32) and (34), where SIĘ is placed after the first conjunct. It would therefore belong to the first conjunct and, under standard coordination rules, could not be shared with the other conjunct.

However, such examples can be accounted for by introducing a general analysis of "intertwined" shared dependents under coordination using the following rules, where DEP corresponds to the shared dependent and RM corresponds to SIĘ:

(59) IPtop → IP DEP Conj IP
↓∈↑ ↑=↓ ↑=↓ ↓∈↑

(60) DEP → { ARG | MOD | RM }
(↑ GF)=↓ ↓ ∈ (↑ ADJUNCT) ↑=↓

Such rules provide an account of examples such as the following ones, which feature plain argument sharing, as well as (32) and (34), which feature sharing of SIĘ:

(61) Przyjechali żandarmi i chodzili od domu
came.PL.M1 military policeman.NOM.PL.M1 and walked.PL.M1 from house
do domu.
to house
'Military policemen came and they walked from house to house.' (NKJP)

(62) Zakleiła kopertę i wepchnęła do torebki.
sealed.SG.F envelope.ACC and pushed into handbag
'She sealed the envelope and she pushed it into the handbag.' (NKJP)

In (61) the shared dependent is the subject *żandarmi*, while in (62) it is the object *kopertę*.[12] In (32) verbs share SIĘ of the same type (reciprocal), while in (34) they share SIĘ of different types (inherent vs. reflexive). The analysis proposed above makes it possible to account for all these examples.

While Frank 2002 discusses German examples similar to (61) and offers an analysis of such coordination in terms of Grammaticalised Discourse Functions (GDF) and structure sharing of relevant f-structure fragments between conjuncts, her analysis does not seem to be applicable in the context of Polish. First, in German this phenomenon is strictly limited to the SUBJ grammatical function, which is not the case in Polish, where other dependents can also be shared, which includes the object, as shown in (62), adjuncts (not presented here for reasons of space) and, more importantly, SIĘ. Secondly, since SIĘ, a weak form that cannot be stressed (Kupść 2000), is under the current analysis a marker rather than an argument (this

[12](62) also features a shared implicit subject.

is uncontroversial for inherent and impersonal SIĘ), it seems inappropriate to assign it discourse functions such as TOPIC or FOCUS. The last option discussed in Frank 2002, namely the SUBJ grammatical function, is not taken into account since SIĘ is not the subject in Polish. Finally, the strongest argument against adopting such an analysis for Polish is that structure-sharing of the f-structure fragment corresponding to SIĘ would make it impossible to account for sentences such as (34), where conjuncts require different basic types of SIĘ – since particular verbs call (45) to specify the type of required SIĘ, two different calls would result in a multifunctional specification of SIĘ, which is an unwanted effect (changing the meaning).

Let us now proceed to the discussion of structures produced by the proposed analysis for Polish: (63) is the f-structure corresponding to (32),[13] where both conjuncts require the same type of SIĘ, namely reciprocal SIĘ:

(63)
$$\left[\left\{\left[\begin{array}{ll}\text{PRED} & \text{`KISS}\langle\boxed{1}\rangle\text{'}\\ \text{SUBJ} & \boxed{1}\left[\begin{array}{ll}\text{PRED} & \text{`PRO'}\\ \text{CASE} & \text{NOM}\\ \text{NUM} & \text{PL}\end{array}\right]\\ \text{SIE} & \left[\begin{array}{ll}\text{RECIP} & +\\ \text{PRESENT} & +\end{array}\right]\end{array}\right], \left[\begin{array}{ll}\text{PRED} & \text{`HUG}\langle\boxed{1}\rangle\text{'}\\ \text{SUBJ} & \boxed{1}\\ \text{SIE} & \left[\begin{array}{ll}\text{RECIP} & +\\ \text{PRESENT} & +\end{array}\right]\end{array}\right]\right\}\right]$$
$$\text{COORD-FORM}\quad \text{AND}$$

The f-structure in (64) corresponds to (34), where coordinated verbs take different types of SIĘ (inherent vs. reflexive). Accounting for the fact that the coordinated predicates in (34) require two different basic types of SIĘ is possible due to the fact that the call to the template SIE defined in (40) (placed in the lexical entry of SIĘ) introduces the defining equation $(\uparrow \text{SIE PRESENT}) = +$, which distributes over all conjuncts, but the constraining equation defined in (43) uses functional uncertainty, which makes it possible to evaluate it separately for each conjunct, which in turn makes it possible for particular conjoined verbs to feature different types of SIĘ – as discussed in § 6.1, it is the verb that specifies the type of SIĘ using the defining equation in (45).

(64)
$$\left[\left\{\left[\begin{array}{ll}\text{PRED} & \text{`LAUGH}\langle\boxed{1}\rangle\text{'}\\ \text{SUBJ} & \boxed{1}\left[\begin{array}{ll}\text{PRED} & \text{`PRO'}\\ \text{CASE} & \text{NOM}\\ \text{GEND} & \text{M1}\\ \text{NUM} & \text{PL}\\ \text{PERS} & 3\end{array}\right]\\ \text{SIE} & \left[\begin{array}{ll}\text{INH} & +\\ \text{PRESENT} & +\end{array}\right]\end{array}\right], \left[\begin{array}{ll}\text{PRED} & \text{`KNOCK}\langle\boxed{1},\boxed{2}\rangle\text{'}\\ \text{SUBJ} & \boxed{1}\\ \text{OBL} & \boxed{2}\left[\begin{array}{ll}\text{PRED} & \text{`HEAD'}\\ \text{PFORM} & \text{W}\\ \text{CASE} & \text{ACC}\\ \text{NUM} & \text{PL}\end{array}\right]\\ \text{SIE} & \left[\begin{array}{ll}\text{REFL} & +\\ \text{PRESENT} & +\end{array}\right]\end{array}\right]\right\}\right]$$
$$\text{COORD-FORM}\quad \text{AND}$$

The proposed analysis also accounts for the impossibility of sharing SIĘ under

[13] As shown in (63) and (64), the implicit subject is shared by the coordinated verbs.

coordination where the predicate in one conjunct would lexically require SIĘ (inherent, reflexive or reciprocal), while with the other SIĘ would be constructional (impersonal), as in the following putative example from Kupść 2000:[14]

(65) Dziecko potykało się i trudno mu *(się)
child.NOM.SG.N stumbled.3.SG.N INH and difficult it.DAT IMPRS
wchodziło po schodach.
climbed.3.SG.N after stairs
'The child stumbled and had difficulty climbing the stairs.'
(Kupść 2000, p. 150, ex. (3.127))

Under the current analysis (65) is ungrammatical not because *Dziecko* is the shared subject (this is because it is not obligatorily shared), which would conflict with the impersonal SIĘ introducing its own implicit subject (as discussed in § 2.4 and formalised in (44)). It is ungrammatical due to the way (40), the template called in the lexical entry of SIĘ, is defined – its second conjunct is a disjunction of two template calls: (41) and (42). The latter, (42), introduces the exclusively impersonal SIĘ, which fails to satisfy the requirements of POTYKAĆ 'stumble' in (65). By contrast, the former, (41), introduces lexically required SIĘ (inherent, reflexive or reciprocal), which can also optionally be impersonal (as a result of multifunctionality allowed in (41)), but it still fails to satisfy the requirements of WCHODZIĆ 'climb' as it does not take any type of lexically required SIĘ – it is exclusively impersonal in (65).

One possible problem for the proposed analysis is, however, how to restrict structure-sharing of the shared dependent such as SIĘ to relevant predicates in examples such as the following one, where only some conjuncts require SIĘ:

(66) Bał się, płakał i śmiał.
feared.SG.M1 INH cried.SG.M1 and laughed.SG.M1
'He was afraid, he cried and he laughed.'

In (66) the first (BAĆ) and the last (ŚMIAĆ) conjunct require the inherent SIĘ, while PŁAKAĆ 'cry' has no such requirement – it cannot therefore satisfy the constraining equations defined in (43). Currently the only alternative is (42), the template introducing impersonal SIĘ, but this would not only distort the semantics of the relevant predicates as SIĘ would be impersonal in all conjuncts, but it would also conflict with the inherent SIĘ specification introduced lexically using (45) by the edge conjuncts (BAĆ and ŚMIAĆ), as (42) requires SIĘ to be only impersonal and blocks multifunctionality (by negating SIE-LEX).

7 Problematic issues: split analysis

The following data concerning the licensing of secondary predicates, in particular the acceptability of (69), is problematic for the uniform analysis of SIĘ as a marker:

[14] Two examples from Kupść 2000, (3.127a)–(3.127b) were merged into one in (65). The glosses and the free translation were modified.

(67) Ja jeszcze nigdy nie widziałem siebie pijanego.
I yet never NEG saw self.GEN drunk.GEN
'I have never seen myself drunk.' (Google)[15]

(68) Ja jeszcze nigdy nie widziałem siebie pijanym.
I yet never NEG saw self.GEN drunk.INST

(69) ?Ja jeszcze nigdy nie widziałem się pijanego/pijanym.
I yet never NEG saw REFL drunk.GEN/INST

In (67)–(68) SIEBIE, corresponding to the object marked for structural genitive case (because sentential negation is present), licenses the secondary predicate PIJANY 'drunk', which may either agree in case (genitive) or appear in the non-agreeing instrumental. On the analysis of SIĘ as a marker, this should not be possible when SIĘ is used instead of SIEBIE, as in (69). However, this example is – perhaps somewhat marginally – acceptable, and multiple acceptable examples of this kind, with the instrumental secondary predicate, may be found in corpora, e.g.:

(70) Widział się gołym.
saw.M1 REFL naked.INST.M1
'He saw himself naked.' (NKJP)

If only the instrumental version were acceptable, then perhaps such examples could be explained away by claiming that the secondary predicate refers to the nominative subject. However, attested examples of secondary predicates apparently agreeing in case with SIĘ may also be found, even if they are much rarer and a little marginal:

(71) ?Nie widzę się zmęczonej, w nowej sytuacji, wśród tabunów babć,
NEG see REFL tired.GEN in new situation among hordes grandmas
cioć itp.
aunts etc
'I can't see myself tired, in a new situation, among hordes of grandmas and aunts, etc.' (NKJP)

Such cases are easy to analyse if SIĘ is treated as a case-bearing argument, but not when it is always claimed to be a marker.

Further evidence against treating SIĘ uniformly as a marker comes from referential properties of this element in verb chains:

(72) Piotrowicz, który z wykształcenia jest mechanikiem,
Piotrowicz.NOM.SG.M1 who from education is mechanic
kazał się tytułować profesorem.
ordered.SG.M1 REFL address.INF professor.INST
'Piotrowicz, who is a car mechanic by profession, ordered people to address him as a professor.' (NKJP)

In (72) SIĘ clearly refers to *Piotrowicz*, which is the subject of the main verb, KAZAĆ 'order'. Because KAZAĆ is an object control verb, its dative object (implicit in (72)) is, as a result of structure sharing, at the same time the subject of the

[15] http://naszeblogi.pl/47575-ujawniam-kolejne-tasmy-rozmowa-tusenko-z-senkiewiczem

embedded infinitival clause headed by TYTUŁOWAĆ 'address'. If, as proposed in § 6, SIĘ is treated as a marker and not as an argument, the predicate TYTUŁOWAĆ can only refer to its own subject, which would block the intended reading of (72). In this example SIĘ behaves in the same way as SIEBIE, which, according to Polish binding rules, typically[16] refers to the subject, but it does not have to be the subject of the predicate local to SIEBIE – it may also be the subject from a higher verb in the verb chain. This ambiguity is illustrated below:

(73) Jan$_i$ kazał Piotrowi$_j$ kupić sobie$_{i/j}$ [...] książkę.
Jan.NOM ordered Piotr.DAT buy.INF self.DAT book.ACC
'Jan ordered Piotr to buy himself a book.'
(Przepiórkowski et al. 2002, p. 178, ex. (6.11a))

(74)
$$\begin{bmatrix} \text{PRED} & \text{'ORDER}\langle\boxed{1},\boxed{2},\boxed{3}\rangle\text{'} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix}\text{PRED} & \text{'JAN'}\end{bmatrix} \\ \text{OBJ}_\theta & \boxed{2}\begin{bmatrix}\text{PRED} & \text{'PIOTR'}\end{bmatrix} \\ \text{XCOMP} & \boxed{3}\begin{bmatrix} \text{PRED} & \text{'BUY}\langle\boxed{2},\boxed{4},\boxed{5}\rangle\text{'} \\ \text{SUBJ} & \boxed{2} \\ \text{OBJ} & \boxed{4}\begin{bmatrix}\text{PRED} & \text{'BOOK'}\end{bmatrix} \\ \text{OBJ}_\theta & \boxed{5}\begin{bmatrix}\text{PRED} & \text{'SELF'}\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

The f-structure corresponding to (73)[17] is provided in (74). As indicated in glosses, (73) has two readings: one where SIEBIE is bound by the local subject, structure shared with the matrix dative object (*Piotrowi*), and another one, where it is bound by the matrix subject (*Jan*).

(75) is similar to (73) in that it also has two readings, depending on which argument binds SIĘ: the intended one where it is the father (the matrix subject) who is to be shaved and another one, somewhat unnatural because the hairdresser is specified for feminine gender, where the hairdresser is ordered to shave herself (the subject of the embedded predicate, local to SIĘ as a result of structure sharing with the matrix object marked for dative case).

(75) Ojciec kazał fryzjerce się ogolić.
father.NOM.SG.M1 ordered hairdresser.DAT.F REFL shave.INF
'Father ordered the hairdresser to shave himself/herself.'

Since the proposed analysis adopts the marker analysis of SIĘ (see § 6), it produces the f-structure in (76) as the representation of (75), which makes the intended reading where it is the father who is to be shaved unavailable, leaving the alternative unintended reading where SIĘ is bound by the local subject, the hairdresser.

[16] SIEBIE is bound by the subject unless it is used with a reciprocal predicate such as in the example below where it is bound by the object:

(i) Przedstawił sobie (nawzajem) sąsiadów.
introduced.SG.M1 self.DAT reciprocally neighbour.ACC.PL
'He introduced the neighbours to each other.'

[17] The glosses and free translation in (73) were added for the purposes of this paper.

(76)
$$\begin{bmatrix} \text{PRED} & \text{`ORDER}\langle \boxed{1},\boxed{2},\boxed{3}\rangle\text{'} \\ \text{SUBJ} & \boxed{1}\begin{bmatrix}\text{PRED} & \text{`FATHER'}\end{bmatrix} \\ \text{OBJ}_\theta & \boxed{2}\begin{bmatrix}\text{PRED} & \text{`HAIRDRESSER'}\end{bmatrix} \\ \text{XCOMP} & \boxed{3}\begin{bmatrix} \text{PRED} & \text{`SHAVE}\langle \boxed{2}\rangle\text{'} \\ \text{SUBJ} & \boxed{2} \\ \text{SIE} & \begin{bmatrix}\text{REFL} & + \\ \text{PRESENT} & +\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

This problem cannot be solved by adopting a clause union analysis of verb chains in Polish, as this would cause problems for the treatment of negation-related phenomena in this environment, which includes handling of genitive of negation and licensing of n-words discussed in Patejuk and Przepiórkowski 2014 in the context of passive. It seems that the only alternative would be to treat reflective (and possibly reciprocal) SIĘ as also having an incarnation as an argument, which would lead to a split analysis of SIĘ. This, however, requires further research.

8 Conclusion

This paper offered an LFG analysis of SIĘ – the Polish so-called reflexive marker – showing its different uses, providing an account of a number of related phenomena, which include multifunctionality of SIĘ, where it has more than one use with respect to the same predicate, haplology of SIĘ in verb chains and finally structure sharing of SIĘ under coordination. The proposed analysis assumes that SIĘ is a marker rather than an argument, but binding facts in verb chains suggest that its status requires further research, which might lead to a split analysis of SIĘ.

Acknowledgements

We are grateful to both reviewers for their comments, which led to various improvements in the form and content of this paper.

Work reported here has been partially financed by the Polish Ministry of Science and Higher Education within the CLARIN ERIC programme 2015–2016 (`http://clarin.eu/`).

References

Dalrymple, Mary, King, Tracy Holloway and Sadler, Louisa. 2009. Indeterminacy by underspecification. *Journal of Linguistics* 45, 31–68.

Dziwirek, Katarzyna. 1990. Default Agreement in Polish. In Katarzyna Dziwirek, Patrick Farrell and Errapel Mejías-Bikandi (eds.), *Grammatical Relations: A*

Cross-Theoretical Perspective, pages 147–161, Stanford, CA: CSLI Publications.

Frank, Anette. 2002. A (Discourse) Functional Analysis of Asymmetric Coordination. In Miriam Butt and Tracy Holloway King (eds.), *The Proceedings of the LFG'02 Conference*, National Technical University of Athens: CSLI Publications.

Jackendoff, Ray. 1992. Madame Tussaud Meets the Binding Theory. *Natural Language and Linguistic Theory* 10, 1–31.

Kibort, Anna. 2004. *Passive and passive-like constructions in English and Polish*. Ph.D. dissertation, University of Cambridge, Cambridge.

Kupść, Anna. 2000. *An HPSG Grammar of Polish Clitics*. Ph.D. dissertation, Institute of Computer Science, Polish Academy of Sciences and Université Paris 7, Warsaw.

Mańczak, Witold. 1956. Ile rodzajów jest w polskim. *Język Polski* XXXVI(2), 116–121.

Patejuk, Agnieszka and Przepiórkowski, Adam. 2012. Towards an LFG parser for Polish: An exercise in parasitic grammar development. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2012*, pages 3849–3852, ELRA, Istanbul, Turkey.

Patejuk, Agnieszka and Przepiórkowski, Adam. 2014. In favour of the raising analysis of passivisation. In Miriam Butt and Tracy Holloway King (eds.), *The Proceedings of the LFG'14 Conference*, pages 461–481, Stanford, CA: CSLI Publications.

Przepiórkowski, Adam, Bańko, Mirosław, Górski, Rafał L. and Lewandowska-Tomaszczyk, Barbara (eds.). 2012. *Narodowy Korpus Języka Polskiego [Eng.: National Corpus of Polish]*. Warsaw: Wydawnictwo Naukowe PWN.

Przepiórkowski, Adam, Kupść, Anna, Marciniak, Małgorzata and Mykowiecka, Agnieszka. 2002. *Formalny opis języka polskiego: Teoria i implementacja*. Warsaw: Akademicka Oficyna Wydawnicza EXIT.

FEATURES AND SELECTION IN LFG: THE ENGLISH VP

John Payne and Kersti Börjars

The University of Manchester

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King (Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

The kinds of analysis that can be provided for selection and agreement phenomena depend significantly on the choices made about the underlying features. In this paper, we review the features that have been used in LFG for the analysis of English verb forms, and propose a motivated alternative which has the consequence that all selection and agreement can be handled through unification.*

1 Introduction

Selection (aka form-government) and agreement are generally construed as being distinct linguistic phenomena. For example, in a well-documented typological survey of the phenomena which might be considered as agreement, Lehmann (1982: 205) argues that these should be "strictly distinguished" from phenomena involving selection, which he claims "arises from government". For Lehmann, selection involves a syntagmatic relation of subordination, where one element (the governor) inherently determines the syntactic function which the subordinate element has in the construction, and hence its correct morphological form.

The forms of the English verb appear to be paradigm illustrations of the intended distinction. First of all, we have one set of forms which have the potential to occur in a subordinate relation to a governor. The governor then determines both their syntactic function and their morphological form. We illustrate here with lexical verbs, but in what follows we will also consider auxiliaries.[1]

(1) a.	helped [**arrest** the thieves]	[infinitive]
b.	kept [**arresting** the thieves]	[gerund-participle]
c.	got [**arrested**]	[past participle]

Here *helped*, *kept* and *got* are governors which require a complement of a particular type, and the head of the complement assumes a particular morphological form. One canonical characteristic of selection is that the form of the subordinate verb does not co-vary with the form of the governing verb. In (1a), for instance, the verb *arrest* takes the infinitive form independently of

* We are grateful for discussion with participants in the Workshop on Morphology held in conjunction with LFG2015 and for comments received from one anonymous reviewer and the two editors of this volume.

[1] The terms we employ for the English verb forms are those of Huddleston (2002a).

the form required for *help*. Cann (1982: 44) indeed makes this property the defining characteristic of selection (which he calls "form-government"). Under both Lehmann's and Cann's conceptions, selection is inherently unidirectional in that one element selects and the other is selected. The selected element does not influence the selector.

Secondly, there is a set of finite forms which have the potential to occur as the predicate of a main clause, that is, in an unsubordinated environment:

(2) a. **arrested** the thieves [past]
b. **arrests** the thieves [3S present]
c. **arrest** the thieves [plain present]

While the past form is invariant, the present forms, plain *arrest* and third-person singular *arrests*, co-vary with forms of the subject. While they differ in some respects, this covariance property is inherent both to Lehmann's (1982: 203) definition of agreement, and to Cann's (1982: 44). See also Steele (1978: 610).

What is not immediately obvious, however, is whether agreement should or should not be considered a unidirectional relation. Probably the majority view falls on the side of unidirectionality. Lehmann (1982: 228), for example, talks of the "source" of agreement, which in the case of subject-verb agreement in English would be the subject NP. For Lehmann's definition of agreement to be satisfied, the subject NP must belong to a person and number category independently of whether there is an agreeing form or not. Cann (1982) talks of "concord controllers" in a similar spirit, and the notion of controllers and targets of agreement is inherent to Corbett's (2006) typology of agreement.

Nevertheless, the issue is not clear-cut. First of all, examples abound in which the supposed targets of agreement are marked with the appropriate agreement categories independently of there being a syntactically overt source for them. This is obviously the case in pro-drop languages, where the subject is not overtly present. Ways can be proposed to circumvent this issue, essentially by proposing a covert source. More interesting though are cases where the subject is syntactically present, but does not belong to one or more of the agreement categories in question. One of the earliest examples is provided by Moravcsik (1978: 351):

(3) Nadie lo vimos [Spanish]
nobody him see.PRT.1P
'Nobody of us has seen him.'

Here the subject NP is clearly not syntactically distinguishable as 1st person, and treating the subject as the source of the person marking would be contrived. Rather, the fact that the subject must be construed as 1st person emerges from the verbal marking. In such cases, what seems to be required is that the forms in the supposed agreement relation are not incompatible with each other, rather than that one of them exclusively acts as the source (see Lehmann 1982: 218-9).

In this paper, we consider the formal mechanisms required to handle both the selection and agreement of English verb forms. One fundamental issue is whether the unidirectionality (or the lack of it) of these relations can be handled by the formalisms in question. A related issue is whether the formalisms need to reflect any unidirectionality which actually obtains. Our impression is that the decisions which have been taken on these points arise first from theoretical considerations internal to the theory involved, and secondly from assumptions concerning the features involved. They tend not to arise from a full perspective on the phenomena involved.

For example, in the Minimalist Program (Chomsky 1995, 1998), a crucial initial distinction is made between interpretable features (those which must survive a derivation and form part of logical form) and uninterpretable features (those which have no semantic content and must therefore be eliminated during the course of the derivation). To this is added the theoretical decision to invoke a mechanism of "checking" that removes uninterpretable features in certain structural configurations (e.g. specifier-head). Since the features which are involved in subject-verb agreement in English (person and number) are deemed to be interpretable and inherent to the subject, the verb must bear identical uninterpretable versions of the same features which are checked during the derivation and eliminated. In other words, the formalisation of person and number agreement is inherently unidirectional because of the a priori decision to distinguish between interpretable and uninterpretable features. This is captured in the terminology associated with feature checking, which treats the verb as the "probe" and the subject as the "goal", i.e. the verb needs to find something to agree with. It is however difficult to reconcile this formalisation with examples such as (3), where the verb alone is the apparent source of the person feature.

On the other hand, in the Minimalist Program a verb and its complement in a selection relationship both bear uninterpretable versions of the relevant feature (e.g. case for NPs, verb-form for VPs). Both features in the feature-checking pair must be eliminated by checking, and selection is therefore essentially formalised without unidirectionality. At first sight, this formalisation appears not to reflect the notional unidirectionality of the selection relation. However, it is not incompatible with it. Essentially, whatever feature is required by a governed constituent can with impunity also be assigned to the governor, and then eliminated.

In this paper, we review the treatment of the selection and agreement of English verb-forms in LFG. In LFG the underlying principle which ensures compatibility of features is functional uniqueness: a feature can have only one value. This principle does not in itself force a unidirectional treatment either of selection or of agreement: all that it requires is compatibility of the features involved, and not necessarily their identity. Examples such as (3) indeed suggest that a formalisation which eschews unidirectionality, i.e. a unificational approach, might be an essential component of any successful approach to subject-verb agreement (even if not of selection). One of the themes we will pursue is

whether the generalisations we wish to capture in the English data in fact point to unification as a solution for selection as well.

2 Selection

2.1 Falk (2001)

Merely as an illustration of an apparently innocuous set of decisions concerning English verbal features within LFG, let us first consider the approach in Falk (2001: 85). A summary of the features and their respective values is given in (4).

(4)

FORM	**FEATURE**	**VALUE**
helps	TENSE	pres
helped	TENSE	past
helping	VFORM PART	pres
helped	VFORM PART	past
help	nothing	

This feature set however raises immediate issues.

Firstly, no feature distinction is made between the finite present form *help* and the infinitive form *help*, while finite past form *helped* is distinguished from past participle *helped*. Whether this might be justifiable depends on the extent the English verbal paradigm can be considered syncretic. The first criterion which is invoked in general discussions of syncretism is whether there are other items which show the relevant distinction. In the case of *helped* (past) and *helped* (past participle), there are a fair number of irregular verbs with distinct past and past participle forms, e.g. *fell* (past) vs. *fallen* (past participle). On this basis, it seems clear that *helped* should be considered syncretic between past and past participle. For these irregular verbs, the distinction is clearly needed for the purpose of selection. For example, we need perfect *have* to select the past participle *fallen* rather than the past *fell*. In the case of infinitive *help* and present non-3SG *help*, only the verb *be* manifests the distinction, with present *am* (1SG)/*are* and infinitive *be*. However, this distinction is also crucial to selection: we need the modal verb *must* to select the infinitive *be* rather than any of the present forms.

In essence, the feature set in (4) treats the additional forms of *be* as cases of "overspecification" (for this term see Baerman 2005, Brown and Hippisley 2012). However, there is a clear difference between the mechanisms

which are needed for the selection of the correct forms of *be* and the mechanisms which are postulated to handle canonical examples of overspecification such as the additional prepositional case forms of a handful of Russian nouns. In Russian, the existence of these nouns forces a single prepositional case feature to be differentiated into two, such that the special additional form is selected by one set of prepositions (e.g. *na* 'on' and *v* 'in') while the regular form is selected by the other (e.g. *o* 'about'). To handle this situation, Brown & Hippisley (2012) invoke a mechanism of default inheritance. That is, there is a feature hierarchy in which a single feature PREP is divided into PREP1 (regular) and PREP2 (additional). Standard nouns which do not make the distinction between PREP1 and PREP2 then simply inherit by default the form appropriate to the undivided PREP feature. Given the feature set in (4), however, a mechanism such as default inheritance could not ensure the selection of the infinitive form of regular verbs such as *help*. This is because there is no feature which the infinitive and the present (non-3S) share to the exclusion of other forms.

Secondly, the choice of a feature set such as (4) entails the use of mechanisms such as constraining equations to handle selection. The appropriateness of constraining equations is clearly stated by Kaplan and Bresnan (1995 [1982]: 63):

> 'A constraining equation is appropriate if, (...), an unspecified value is intended to be in conflict with all of a feature's real values. On the other hand, a value specification may be omitted for some features as an indication of vagueness and the restriction is then naturally stated in terms of a defining equation.'

For example, a defining equation such as (5) will correctly ensure that the gerund-participle form of *hand* is selected by progressive *be* in (6a), and rule out the selection of the past participle *handed* as in (6b):

(5) *be*: (↑VCOMP PART)= *pres*

(6) a.. The girl is handing the baby a toy.
b. *The girl is handed the baby a toy.

This is because *handed* can be assigned PART=*past* via the equation (↑VCOMP PART)= *past* associated with the verb *be*. However, under Falk's analysis (5) does not rule out (7a), because *hands* has been assigned no value for the feature PART, or (7b), because *hand* has no features at all.

(7) a. *The girl is hands the baby a toy.
b. *The girl is hand the baby a toy.

Changing (5) to the constraining equation (8) solves this issue.

(8) *be*: (↑VCOMP PART)= c *pres*

Now only (6a) is indeed permitted.[2]

The introduction of constraining equations such as (8) brings an essentially unidirectional mechanism into play, since the item which bears the constraining equation determines the form of its dependent rather than vice versa. No descriptive difficulties arise from this as far as the selection of English verb forms is concerned. From a mathematical perspective, however, there is a potential cost. In an attempt to formalize the LFG construction algorithm declaratively, Blackburn and Gardent (1995: 44) note that constraining equations, which effectively test between a multiplicity of already generated representations, constitute "a dynamic residue that resists a purely declarative analysis". See also Hancox (1994, 2003) and Börjars and Payne (2013). The fact that constraining equations are necessary appears to follow directly from the initial choices that are made concerning features, rather than from any inherent necessity in the theory.

2.2 An alternative

In this section we propose an alternative set of features for English verb forms which is intended in the first instance to capture what we believe to be genuine generalisations, but also has the consequence that essentially unidirectional mechanisms such as constraining equations prove to be unnecessary.

The proposed alternative is given in (9), PERS is not included here, we will return to a discussion of this feature in Section 3:

FORM	**FEATURE**	**VALUE**
helps	VFORM	*fin*
	TENSE	*prs*
	NUM	*sg*
help	VFORM	*fin*
	TENSE	*prs*
	NUM	*nonsg*

[2] Although the feature set used is different for non-finite forms, the Pargram starter grammar for English employs a similar VFORM feature. See: http://www2.parc.com/isl/groups/nltt/xle/doc/PargramStarterGrammar/eng-pargram-lex.lfg. One essential difference between this and the proposal made here is that the Pargram grammar uses constraining equations in its CHECK algorithms for selection by English auxiliaries. A second is that auxiliaries do not head their own clauses (see below).

(9)

helped	VFORM TENSE	*fin* *pst*
helping	VFORM	*ger-part*
helped	VFORM	*pst-prt*
help	VFORM	*infin*

Some aspects of (9) are immediately apparent, notably the fact that present and infinitive forms are not treated as syncretic, and that all verb forms, not just participles, should have a VFORM feature. The decision not to treat present and infinitive forms as syncretic is justified by the behavior of *be*, and the decision to allocate the VFORM feature to all forms removes an obvious anomaly.

These two decisions immediately remove any necessity to use constraining equations for the purposes of selection. To be specific, the defining equation in (10) will correctly permit (11a) and disallow (11b-d).

(10) *be*: (↑XCOMP VFORM)= *ger-prt*

(11) a.. The girl is handing the baby a toy.
b. *The girl is handed the baby a toy.
c. *The girl is hands the baby a toy.
d. *The girl is hand the baby a toy.

The form *handed* has either the value *fin* or *pst-prt* for VFORM, and the form *hands* has either *fin* or *infin*.

A further characteristic worth noting is that there is a class of finite forms. This can be justified by the fact that finite forms essentially have the same distribution. For example, they require nominative subjects and may enter into agreement relations with their subjects. The presence of a finite verb form also plays a role in selection by main verbs. For example, the verb *think* requires a finite verb form to head its complement in (12b), and this is captured by the defining equation in (12b).

(12) *think*: (↑COMP VFORM)= *fin*

(13) a. I think [$_{IP}$ this is a good start].
b. I think [$_{IP}$ this was a good start].

Note that the subordinate clause in (13) is an IP, and that the choice of verb form cannot be made contingent on the presence of a complementizer.

What is not the case in (9) is that there is claimed to be a class of non-finite forms. This is justified by the fact that, unlike the finite forms, the participles and the infinitive form do not have similar distributions and are selected distinctively by different verbs. Falk (2001) suggests that a

generalization linking the participles might be their ability to occur as pre-head adjuncts in noun-phrase structure, as in (14).

(14) a. the winning team [gerund-participle]
b. the defeated army [passive past participle]

However, this apparent generalization does not survive close scrutiny. Firstly, it is not the case that exclusively the participles can function as nominal adjuncts. As can be seen in (15a), infinitives also occur in this environment:

(15) a the soon to be released prisoners [infinitive]
b. ??the to be released prisoners

Of course, the occurrence of infinitives in this function is severely constrained: the infinitive has to be modified by an adverb such as *soon* (compare 15b). However, this property does not distinguish the infinitive as pre-head modifier from all participles. As shown in (16), a similar constraint applies in the case of active past participles (but involving a different set of adverbs).

(16) a. the newly arrived guests [active past participle]
b. ??the arrived guests

These complexities suggest that there is in reality no overall generalisation concerning the use of verb forms as pre-head modifiers. Each of the non-finite forms has its own restrictions and peculiarities.

To conclude this section, we note that our treatment of selection implies that auxiliary verbs in English select their complements in a similar way to main verbs. In other words, we adopt an "auxiliaries as main verbs" analysis (Pullum and Wilson 1977, see also Dyvik 1999) rather than a Pargram-style approach in which auxiliary-plus-verb constructions are treated as monoclausal. This has the cost that English auxiliary-plus-verb constructions cannot be treated in a parallel manner to those of languages such as French (as in Butt et al. 1999). However, there is a body of independent evidence in English that auxiliaries do indeed head their own clauses. These include the ability of the auxiliary and complement to have their own independent temporal specification and negation (for a synopsis see Huddleston 2002b: 1209ff). As an illustration of the independent negation property, even with an auxiliary such as dummy *do*, see (17):

(17) I didn't not phone you [Sliding Doors, 1998]

This example, spoken by the hero in a well-known film, is clearly intended not simply to communicate that the phone call was actually made. The subordinate negation in this case seems like antonymic rather than classical negation: i.e. from the negation of "not phoning" we are not allowed to conclude "phoned".

See Hamm and Lambalgen (2005, citing E. Engdahl as source) for an example of antonymic negation of the complement of a perception verb, i.e. a main verb:

(18) The policeman saw Andrew not stop at the lights

This analysis also implies that a semantic account must be provided of restrictions on auxiliary order (for a review of how this might be done see Falk 1984).

3 Agreement

3.1 Existing accounts

Existing accounts of subject-verb agreement in English all assume in some form a basic distinction of singular versus plural for number, and 1 vs. 2 vs. 3 for person. For the verb *be*, and all other verbs including main verbs, this results in the present tense paradigms in (19):

(19)

FORM		**PERS**	**NUM**
help	am	1	sg
help	are	2	sg
helps	is	3	sg
help	are	1	pl
help	are	2	pl
help	are	3	pl

The main issue then becomes how to deal with the apparent syncretism of the non-3SG forms of main verbs like *help*.

If the form *helps* is taken as representing [PERS *3*, NUM *sg*], and the form *help* is unspecified for person and number, as in (4), then we are forced into the use of constraining equations for subject-verb agreement. See Bresnan (2001: 60) for such an approach. Essentially, the subject-verb agreement rule will have to be as in (20). Some form of morphological blocking will be assumed to favour the more specific form and hence rule out *help* occurring with a third person singular subject.

(20) *helps*: (↑SUBJ PERS) =c *3*
(↑SUBJ NUM) =c *sg*

This rule is unidirectional in nature, and treats the subject as the source of agreement. That however a unificational approach to subject-verb agreement might be preferable can be seen from examples such as (21):

(21) a. What has/have pointed heads and long tails?
b. The sheep has/have not been counted.

For further examples of this kind, see Payne and Huddleston (2002). In these examples, it appears reasonable to treat the subject as unspecified for NUM, and to treat its singularity or plurality as inherited from the verb. These are the analogues of the Spanish example in (3), though with respect to number rather than person.

A unificational approach is of course entirely possible if all present tense forms of English verbs (and not just *be*) are assigned values for the person and number features. However, this is a blunt approach to the syncretisms involved. In sections 3.2 and 3.3 we provide a novel alternative account involving a re-evaluation of person and number features.

3.2 An alternative: the nature of NUM

In our approach, the basic lexical entry for nouns will contain a NUM distinction between *sg* and *nonsg*. Person will not be treated as inherent to nouns, and they will not contain a PERS feature (for motivation of this assumption, see for example Lehmann 1982). This will have the immediate consequence that standard nouns can unify with the first and second person determiners, as in (22).

(22) a. we students
b. you students

An account of the apparent number restriction involved here (first person *I* and singular *you* are not permitted) will be given below.

Collective nouns such as *committee* have an agreement pattern which looks more semantically based. To handle these nouns, we adopt Wechsler's (2003) AGG(REGATE) feature (though as an f-structure feature rather than purely semantic one). By default, [NUM *sg*] implies [AGG –] and [NUM *nonsg*] implies [AGG +]. But bipartite nouns like *scissors* can be interpreted either as singular or plural, even though they obligatorily require plural agreement, and these will be unspecified for AGG. In this system, NUM is essentially a feature which controls syntactic agreement, and AGG is the feature which feeds into the semantics. The features associated with a sample set of nouns is then as in (23):

(23) a. *student* [NUM *sg*, AGG −]

b. *students* [NUM *nonsg*, AGG +]

c. *committee* [NUM *sg*]

d. *scissors* [NUM *nonsg*]

The corresponding verb agreement rule will straightforwardly allow [NUM *sg*] noun phrases to unify with *prs sg* (as well as all *past*) verb forms. On the other hand, the "plural" verb form will be allowed to unify not only with [NUM *nonsg*], but also with [AGG +] noun phrases. This disjunction is inherent already to the double agreement pattern of collective nouns. For a bipartite noun such as *scissors*, the agreement can only be plural, however.

3.3 The nature of PERS

The person feature should be a set-valued feature (Dalrymple and Kaplan 2000) in order to account for coordination. In our treatment, only 1st and 2nd person pronouns and agreeing forms of *be* will be specified for PERS. Following Wechsler (2003), the "plural" forms will all be [AGG +]. "Singular" *you* is [AGG –], but like "plural" *you* regularly has *nonsg* verb agreement, e.g. *help*, *are*. The only unusual form is then *I*, which must likewise be treated as having ***nonsg*** verb agreement, but requiring special forms for *be*. The person and number features associated with the personal pronouns and determiners are then as in (24):

(24) a. *I* [PERS {1}, NUM *nonsg*, AGG –]
b. *you* [PERS {2}, NUM *nonsg*]
c. *he*, *she*, *it* [NUM *sg*, AGG −]
d. *we* [PERS {1}, NUM *nonsg*, AGG +]
e. *they* [NUM *nonsg*, AGG +]

Because of the values associated with NUM, these feature specifications straightforwardly allow all noun phrases containing personal pronouns or determiners to unify straightforwardly with standard verbs. The syncretism seen in standard verbs follows simply from the fact that these have no PERS feature. The present tense forms of standard verbs have only NUM and AGG features, with the appropriate defining equations as in (25).

(25) a. *helps*: (↑VFORM) = *fin*
(↑TENSE) = *prs*
(↑SUBJ NUM) = *sg*

b. *help*: (↑VFORM) = *fin*
(↑TENSE) = *prs*
(↑SUBJ NUM) = *nonsg* ∨ (↑SUBJ AGG) = +

The forms of *be* will be essentially the same, as in (26).

(26) a. *is* (↑VFORM) = *fin*
(↑TENSE) = *prs*
(↑SUBJ NUM) = *sg*

b. *are*: (↑VFORM) = *fin*
(↑TENSE) = *prs*
(↑SUBJ NUM) = *nonsg* ∨ (↑SUBJ AGG) = +

c. *am*: (↑VFORM) = *fin*
(↑TENSE) = *prs*
(↑SUBJ NUM) = *nonsg*
(↑SUBJ AGG) = –
(↑SUBJ PERS) = {1}

Rather than using a more complex set of assigning equations in (26b), we assume that the form *are* can be blocked from unifying with *I* because of the existence of the extra highly marked form *am*, which is the only verb form specified for person. The form *are* does of course surface in the negative interrogative: *aren't I*? However, we leave a fuller discussion of this issue to future work.

The feature specifications for personal determiners have the fringe benefit of predicting the mysterious distinction between (27a) and (27b):

(27) a. *I student will not accept these changes!
b. We students will not accept these changes!

This contrast now follows from the incompatibility of the NUM values for the personal determiner I, which is [NUM *nonsg*], and the noun *student*, which is [NUM *sg*]. The singular **you student* is analogously blocked.

4 Conclusion

In this paper we have proposed a new feature system to account for the selection and agreement properties of English verb forms. The generalisations which these features encapsulate enable both selection and subject agreement to be handled by unification, rather than by unidirectional mechanisms such as constraining equations. In the case of subject agreement, a unification procedure indeed seems to be forced by the data.

Perhaps the most distinctive aspect of this new feature set is the treatment of number. Essentially, our proposal treats person as solely a property of the first and second person pronouns and determiners, whose singular/plural distinction, as in Wechsler (2003), is captured by the feature AGG. This is justified by the well-known fact that the plural of 1st and 2nd person pronouns is not straightforwardly plural: *we* is not 'I + plural', in the sense that when I refer to 'we' I am not referring to a set of 'Is'. The basic number distinction between [NUM *sg*] and [NUM *pl*] is then primarily a characteristic of traditional third person singular and plural forms. Such a feature system allows of course a paradigm of verb forms in which all traditional persons and numbers are expressed differently. But it also permits, as with English standard verbs, a system in which the traditional third person singular is distinguished by being the only form which is [NUM *sg*]. Typologically, such a system seems entirely plausible: the basic distinction in the person-number paradigm, if it exists, is between 3rd person singular and all other forms.

References

Baerman, Matthew 2005. Typology and the formal modelling of syncretism. *Yearbook of Morphology* 2004: 41-72.

Blackburn, Patrick & Claire Gardent 1995. A specification language for Lexical Functional Grammars. In: *Proceedings of the Seventh Conference of the European Chapter of the Association for Computational Linguistics*, 39–44.

Brown, Dunstan and Andrew Hippisley 2012. *Network morphology: a defaults-based theory of word structure*. Cambridge: Cambridge University Press.

Butt, Miriam, Tracy Holloway King, María-Eugenia Niño, and Frédérique Segond 1999. *A grammar-writer's cookbook*. Stanford, CA: CSLI Publications.

Chomsky, Noam 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, Noam 1998. Minimalist inquiries: The framework. In: *MIT Occasional Papers in Linguistics* 15. Cambridge, MA: MITWPL.

Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.

Dalrymple, Mary and Ronald M Kaplan 2000. Feature indeterminacy and feature resolution. *Language* 76: 759–798.

Dyvik, Helge 1999. The universality of f-structure: discovery or stipulation? In Miriam Butt and Tracy Holloway Kind (eds.), *Proceedings of the LFG99 Conference*. Stanford, CA: CSLI Publications.

Falk, Yehuda 1984. The English auxiliary system: a lexical-functional analysis. *Language* 60: 483-509.

Falk, Yehuda 2001. *Lexical-Functional Grammar*. Stanford, CA: CSLI Publications.

Hamm, Fritz and Michiel van Lambalgen 2005. Formal foundations for semantic theories of nominalization. Berlin: ZAS papers

Hancox, Peter 1994. The uniform treatment of constraints, coherency and completeness in a Lexical Functional Grammar compiler. In *AI and*

cognitive science '94: proceedings of the 7th annual conference. Dublin: Dublin University Press, 129–142.

Hancox, Peter 2003. Lexical Functional Grammar constraints and concurrent constraint programming. In *AI and cognitive science '05: proceedings of the 16th annual conference*. Coleraine: University of Ulster, 309–318.

Huddleston, Rodney 2002a. The verb. In: Rodney Huddleston and Geoffrey K Pullum et al., *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press, 71–212.

Huddleston, Rodney 2002b. Non-finite and verbless clauses. In: Rodney Huddleston and Geoffrey K Pullum et al., *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press, 1171–1274.

Kaplan, Ronald M and Joan Bresnan 1982. Lexical-Functional Grammar: a formal system for grammatical representation. In Joan Bresnan (ed.), *The mental representation of grammatical relations*. Cambridge, MA: MIT Press, 173–281.

Kaplan, Ronald M and Joan Bresnan 1995 [1982]. Lexical-Functional Grammar: a formal system for grammatical representation. In Mary Dalrymple, Ronald M Kaplan, John T Maxwell III and Annie Zaenen (eds.), *Formal issues in Lexical Functional Grammar*. Stanford, CA: CSLI Publications, 29–130 [Reprint of Kaplan & Bresnan 1982].

Lehmann, Christian 1982. Universal and typological aspects of agreement. In Hansjakob Seiler and Franz J. Stachowiak (eds.), *Apprehension: das sprachliche Erfassen von Gegenständen* II. Tübingen: Narr, 201– 267.

Moravcsik, Edith A 1978. Agreement. In Joseph H Greenberg, Charles A Ferguson and Edith A Moravcsik (eds), *Universals of human language. Vol IV Syntax*. Stanford: Stanford University Press, 331–374.

Payne, John and Rodney Huddleston 2002. The noun phrase. In Rodney Huddleston and Geoffrey K Pullum et al., *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press, 323–523.

Pullum, Geoffrey K., and Deirdre Wilson 1977. Autonomous syntax and the analysis of auxiliaries. *Language* 53: 741–788

Steele, Susan 1978. Word order variation: a typological study In: Joseph H. Greenberg, Charles A. Ferguson and Edith A. Moravcsik (eds), *Universals of Human Language: IV: Syntax*. Stanford: Stanford University Press, 585-623.

Wechsler, Stephen 2003. Number as person. In Olivier Bonami & Patricia Cabredo Hofherr (eds), *Empirical issues in formal syntax and semantics. papers from CSSP 2003*. Paris: Colloque de Syntaxe et Sémantique á Paris.

AN LFG APPROACH TO NESTED DEPENDENCIES IN DUTCH

Marjolein Poortvliet
University of Oxford

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King (Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

This paper argues that, contrary to the prevailing idea that Dutch is a cross-serial dependency language, Dutch also uses nested dependency constructions in embedded clauses. In these nested dependency constructions, the order of the verb cluster is reversed compared to the order of the noun cluster, much like the order of German embedded clauses. I show that the Dutch nested dependency construction is restricted to passive constructions with perception verbs and causative verbs as the main predicate. I propose an LFG treatment for these restricted nested dependency structures.

1 Introduction

A unique syntactic phenomenon that is typical of Dutch and German is the serial dependency structure.[1] This structure occurs in embedded clauses, in which all verbs appear clause-finally and are preceded by all the verbs' arguments. Unlike the English embedded clauses, which consist of sequential argument-verb pairs (see (1a)), Dutch and German embedded clauses are such that all nominal material comes first, followed by all verbal material. The sentences in (1) provide a comparison of English (see (1a)), Dutch (see (1b)) and German (see (1c)), in which the arrows are used to show the dependency relations between the arguments and their verbs.

1. (a) ...that Peter saw Marie swim (English)

   (b) *...dat Peter Marie zag zwemmen* (Dutch)
   ...that Peter Marie saw swim

   (c) *...dass Peter Marie schwimmen sah* (German)
   ...that Peter Marie swim saw

Dutch is said to be a cross-serial dependency language, demanding the same internal left-to-right order both in the noun cluster and in the verb cluster. German, on the other hand, is said to be a nested-dependency language, in which the order of the verbs in the verb cluster is reversed compared to the order of the arguments in the noun cluster. Even though these claims about Dutch and German still stand in

[†]For helpful comments and suggestions, I thank Mary Dalrymple, Ash Asudeh, Carryn Yong, the audience at LFG15, and the anonymous reviewers.

[1]The terms 'Dutch' and 'German' are used to denote Standard Dutch and Standard German respectively. Other varieties of these languages are also known to have serial dependency structures. For instance, a variety that has raised interest for its serial dependencies is Swiss German, see Zaenen and Kaplan (1995) and Shieber (1985).

most cases, certain highly constrained constructions in Dutch suggest that Dutch also has nested dependencies. Such a construction is shown in (2), in which the past participle of the verb cluster is not in the expected clause-final position as given in (3), but appears before the main predicate *zag* 'saw'.

2. *...dat Jan een lammetje gevoed zag worden*
   ...that Jan a lamb fed saw to-become
   '...that Jan saw a lamb being fed"

3. **...dat Jan een lammetje zag worden gevoed*
   ...that Jan a lamb saw to-become fed

Serial dependency structures have received much attention in the literature from various perspectives: they have been discussed within syntactic frameworks such as Lexical Functional Grammar (e.g. Bresnan et al. 1982, Zaenen and Kaplan 1995, Kaplan and Zaenen 2003), Head-driven Phrase Structure Grammar (e.g. Rentier 1994 and Hinrichs and Nakazawa 1998) and Tree Adjoining Grammar (e.g. Rambow 1992), as well as from psychological perspectives (e.g. Kaan and Vasic 2004) and from comparative perspectives (e.g. Bach et al. 1986). Furthermore, cross-serial dependencies have played an important role in discussions of language complexity and have functioned as evidence for the argument that natural languages are not context-free (see Shieber 1985 for evidence from Swiss German, and Bresnan et al. 1982 for evidence from Dutch).

In this paper, I show how nested dependency structures in Dutch can be treated within the framework of Lexical Functional Grammar by making a number of modifications to the rules proposed for crossed dependencies in Dutch by Kaplan and Zaenen (2003) and Zaenen and Kaplan (1995). The outline is as follows: first, previous research on cross-serial dependency constructions will be revisited, followed by an overview of the rules proposed for these constructions. Secondly, I will introduce a passive construction (see (2)) which cannot be accounted for by the rules proposed and which therefore demand modifications to the existing rules. Finally, I show that this construction is not only limited to past participles, but can also include resultative adjectives and resultative PPs.

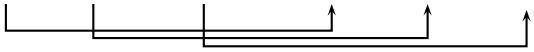
2 Background

2.1 Crossed Dependencies in Dutch

A cross-serial dependency construction consists of a noun cluster and a verb cluster. The noun cluster is formed by the arguments of the main predicate and the arguments of the complement or complements of the main predicate. The noun cluster is ordered according to the embeddedness of the complements, i.e. N1 N2 N3. This noun cluster is followed by a sentence-final verb cluster, formed by the main predicate and the verbs of the complement. The order of the verb cluster

is also ordered according to the embeddedness of the complements, i.e. V1 V2 V3. This means that both clusters have the same relative order of elements. If we were to draw dependency lines between the arguments and verbs, these lines would cross. Therefore, an overarching term for these constructions is *cross-serial dependency structures*. An often cited example (e.g. Seuren and Kempen 2003, Kaplan and Zaenen 2003, and a slightly modified version in Zaenen and Kaplan 1995) of such a sentence is given in (4), in which the matrix verb *wil* 'want' is linked to the subject *Jan*, followed by the complement consisting of the verb *laten* 'let' linked to its object *Marie*, followed by a third complement consisting of the verb *studeren* 'study' linked to its object *geneeskunde* 'medicine'.

4. *...dat Jan Marie geneeskunde wil laten studeren*
   ...that Jan Marie medicine wants to-let to-study

   '...that Jan wants to let Marie study medicine'

There are two types of cross-serial dependency structures in Dutch: *the verb raising construction* as first noted by Evers (1975), and *the third construction* as coined by Besten et al. (1988). A number of verbs in Dutch can appear in both constructions, as is shown for the verb *trachten* 'try' in (5) and (6). In both examples, the infinitive verb *helpen* 'help' from the complement appears to the right of the matrix verbs *heeft trachten* or *heeft getracht* 'has tried', creating the verb cluster V1 V2. Similarly, the argument of the complement, *Marie*, appears to the right of the argument of the matrix verb, creating the noun cluster N1 N2.

5. Verb Raising

   ...dat Jan Marie heeft trachten te helpen
   ...that Jan Marie has try.INF to help

   '...that Jan has tried to help Marie'

6. Third Construction

   ...dat Jan Marie heeft getracht te helpen
   ...that Jan Marie has try.PART to help

   '...that Jan has tried to help Marie'

There are two differences between the verb raising construction and the third construction. First of all, the auxiliary in the verb raising construction selects an infinitive instead of the expected past participle that is found in the third construction. This phenomenon is called the IPP-effect (*Infinitivus pro Participio*) and only occurs if the verb selected by the auxiliary takes a complement itself. The IPP-effect

only takes place in combination with certain verbs, i.e. causatives, modals, perception verbs, duratives, and is optional for inchoatives and control verbs. Secondly, the verb raising constructions can employ both subject and object control relations, but the third construction can only have a subject control verb.

2.2 Nested Dependencies in German

The phenomenon of serial dependencies is also found in German (see Vogel et al. 1996, Bach et al. 1986, for the third construction in German see Rambow 1992, Hinrichs and Nakazawa 1998), but differs from Dutch in that the order of the verbs within the verb cluster in German is different from the order in Dutch: whereas in Dutch a lower verb follows its matrix verb, in German a lower verb precedes its matrix verb. When dependency lines are drawn between the verbs and their arguments in the German sentence, this shows that the dependencies are nested (nouns ordered N1 N2 N3, while verbs are ordered V3 V2 V1). This type of structure is therefore called a nested dependency structure.

7. *...dass Jan Marie Medizin studieren lassen will*
   ...that Jan Marie medicin to-study to-let to-want

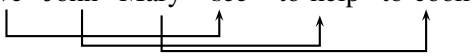
   '...that Jan wants to let Marie study medicin'

2.3 Kaplan and Zaenen's LFG analysis for Dutch

Dutch dependency structures have received a considerable amount of attention within the framework of lexical-functional grammar, and after its first treatment by Bresnan et al. (1982) further research was carried out by Zaenen and Kaplan (1995) and Kaplan and Zaenen (2003). This section provides an overview of the LFG treatment and the rules proposed for cross-serial dependencies in Dutch (leaving the third construction out of the discussion), and is mainly based on the most recent work by Kaplan and Zaenen (2003).

According to the c-structure and f-structure proposed by Kaplan and Zaenen (2003), a sentence such as (8) has the c-structure and the f-structure shown in Figure 1. The c-structure shows that the V node dominating *helpen* 'help', its mother node V'_1 and the NP node *Jan* combined make up the SUBJ and PRED of $XCOMP_1$, whereas the V node dominating *koken* 'cook', its mother node V'_2 and the NP *Marie* make up the SUBJ and PRED of $XCOMP_2$.

8. *...dat we Jan Marie zien helpen koken*
   ...that we John Mary see to-help to-cook

   '...that we see John help Mary to cook'

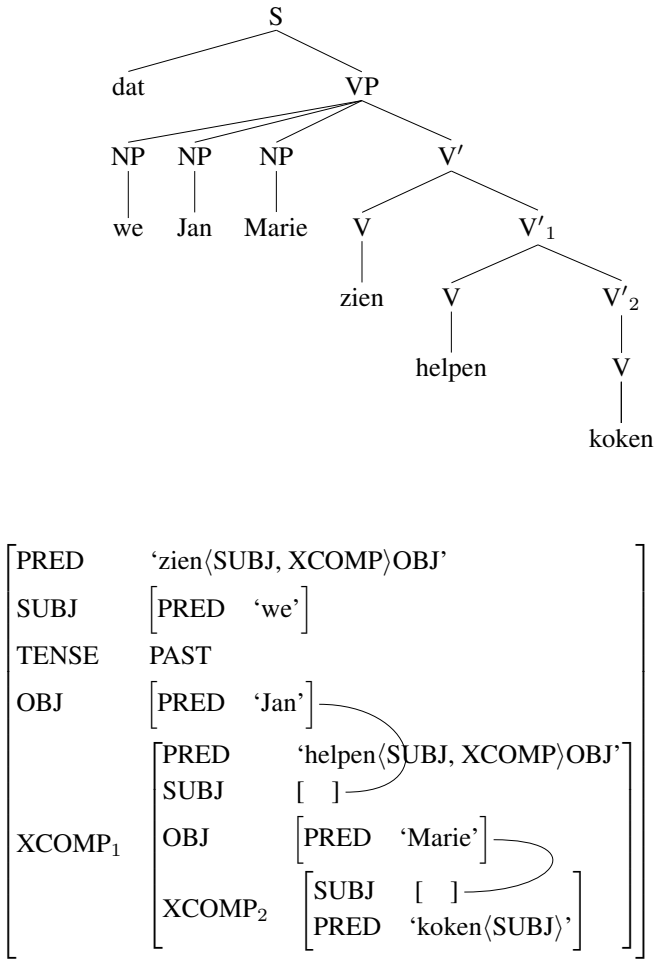


Figure 1: C-structure and f-structure for (8)

The phrase structure rules for Dutch dependency structures as proposed by Kaplan and Zaenen (2003) are given in (5) below. These rules account for the various complementation patterns found in Dutch. Before explaining the technicalities of the equations in (5), I will briefly discuss these complementation patterns. As a starting point, let us take a look at (9), which consists of a simple construction without a complement. The c-structure and f-structure of such a construction without a complement is given in Figure 2, which shows that the NP in the c-structure corresponds to SUBJ at f-structure and that the V′ corresponds to the entire f-structure.

9. *...dat Jan helpt*
   ...that Jan helps
   ‘...that Jan helps’

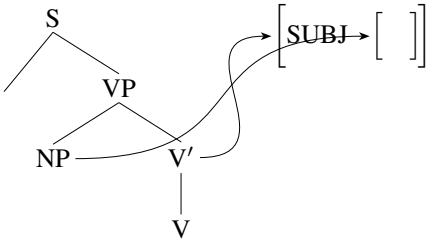


Figure 2: Constructions with matrix verb only

A slightly more complex embedded clause with one complement is given in (10). Such a construction, containing an XCOMP, would have the partial c-structure and f-structure given in Figure 3.[2] For Figure 3 shows the correspondences that were given above in Figure 2, but with the additional correspondences of the second NP to OBJ of XCOMP at f-structure, and the second V′ to XCOMP. More XCOMPs can be added and would be added in the exact same way.

10. *...dat Jan Marie probeert te helpen*
    ...that Jan Marie tries to help
    '...that Jan tries to help Marie'

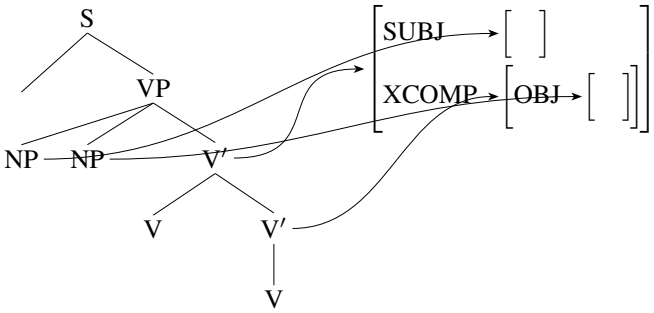


Figure 3: Constructions with matrix verb and one XCOMP

Furthermore, Dutch has the option of extraposing complements within an embedded clause. This is shown in (11), in which the complement *Marie te helpen* is extraposed to the right. Note the difference in position of the NP *Marie* between (11) and (10). The partial c-structure and f-structure of such a construction is given in Figure 4.[3] The f-structure appears very similar to the f-structure in Figure 3, but note that the complement is an COMP instead of an XCOMP, following Zaenen and Kaplan (1995) and Schuurman (1991) for Dutch and Berman (2000) for German.[4]

[2]For the sake of simplicity, the XCOMP's SUBJ is not included in the f-structure, but it is assumed that the main predicate shares its SUBJ with its XCOMP.

[3]Again, the COMP's SUBJ is not included in the f-structure, but based on Zaenen and Kaplan (1995)'s analysis, a 'pro' value for the PRED of the COMP's SUBJ is assumed.

[4]Zaenen and Kaplan (1995) take as evidence for the analysis of the extraposed complement as a

11. *...dat Jan probeert Marie te helpen*
...that Jan tries Marie to help
'...that Jan tries to help Marie'

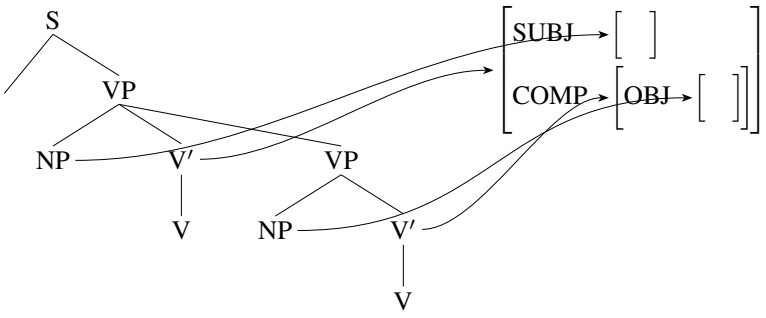


Figure 4: Extraposed constructions

From what has been discussed so far, it is clear that all arguments must precede all verbs, and that both the argument cluster and the verb cluster have specific orders within their cluster. Kaplan and Zaenen (2003), further building on Zaenen and Kaplan (1995), propose the phrase structure rules given in Figure 5. These rules state that a VP can expand to zero or more (as indicated by the Kleene star) NPs, which can serve as some nominal grammatical function (as indicated by NGF) of the complementizer phrases. The functional uncertainty equation (↑XCOMP * (COMP) NGF)=↓ indicates that each grammatical function can be reached in the f-structure by an indeterminate number of XCOMPs. This is followed by a V′ and an optional VP for extraposed structures. Furthermore, the rule in Figure 5 shows that V′ always expands to a V and optionally to an additional V′. The shuffle operator ',' indicates that both the orders V, V′ and V′, V are possible, in order to ensure that both constructions in (12) are permitted.

12. *...dat Jan Marie heeft gezien/gezien heeft*
...that Jan Marie has seen/seen has
'...that Jan has seen Marie'

COMP the fact that impersonal passives are possible with extraposed constructions (i), but not with verb raising constructions (ii):

i ...omdat er (door iedereen) werd getracht Marie te helpen

ii ∗...omdat er (door iedereen) Marie werd trachten te helpen

The functional control equation that is required by the XCOMP function can only specify the subject or object of the higher predicate. Since the XCOMP's SUBJ in the raising construction in (ii) is identified with an oblique agent function (i.e. *door iedereen*), this construction is ungrammatical. The COMP function does not have such a requirement. Therefore, an analysis of extraposed constructions taking COMPs instead of XCOMPs would explain why the impersonal passive is permitted in the extraposed construction in (i).

VP	→	NP*		V′		(VP)	
		(↑ XCOMP* (COMP) NGF)=↓				(↑ XCOMP* COMP)=↓	

V′	→	(PostP)	(V)		[V ,		(V′)]
		(↑ XCOMP$^+$ PRT)=↓	(↓ VFORM)=PART				(↑ XCOMP)=↓
			(↑ XCOMP$^+$)=↓				(↑ XCOMP$^+$ NGF) $\neg <_f$ (↑ NGF)
			↓ $\neg <_f$ (↓ PRT)				

Figure 5: Phrase structure rules for VP and V′

Zaenen and Kaplan (1995) and Kaplan and Zaenen (2003) propose a flat structure of NPs, the embeddedness of which is reflected in the f-structure. The correlation between the level of embedding of the NP and that of its verb is provided by the functional notation below V′ in Figure 5. These annotations under the V′ state that the next lower verb becomes the head of the XCOMP assigned at each level. In order to constrain the cross-serial dependency relations between the NPs and Vs, a rule of f-precedence (i.e. $\uparrow$ XCOMP^{+} NGF) $\neg <_{f}$ ($\uparrow$ NGF) has been added to this rule, which states that an NP that belongs to a lower verb cannot precede an NP belonging to a higher verb.

The main additions to Zaenen and Kaplan (1995) in Kaplan and Zaenen (2003) are the optional PostP and V (as shown in Figure 5) to account for leftward creepers: Dutch has leftward creeping particles (as in (13)) and leftward creeping past participles (see (14)). The extension PostP permits particles of separable verbs to creep leftwards and to assign the feature PRT to the f-structure of their head, which is always at least one level further embedded than the PostP itself, hence the Kleene Plus. The other extension is the optional V when it is a participle, which maps onto an f-structure that is the XCOMP of its mother node. Again, this XCOMP is always at least one level further embedded than the past participle itself. The f-precedence constraint simply states that the past participle cannot come before its particle.

13. (a) *...dat Jan het liedje zal hebben* ***mee****gezongen*
...that Jan the song will have along.sung
'...that Jan will have sung along with the song'

(b) *...dat Jan het liedje zal* ***mee*** *hebben gezongen*
...that Jan the song will along have sung
'...that Jan will have sung along with the song'

(c) *...dat Jan het liedje* ***mee*** *zal hebben gezongen*
...that Jan the song along will have sung
'...that Jan will have sung along with the song'

14. (a) *...dat Jan het liedje zal hebben* ***gezongen***
...that Jan the song will have sung'
'...that Jan will have sung the song'

(b) *...dat Jan het liedje zal* ***gezongen*** *hebben*
...that Jan the song will sung have
'...that Jan will have sung the song'

(c) *...dat Jan het liedje* ***gezongen*** *zal hebben*
...that Jan the song sung will have
'...that Jan will have sung the song'

3 Nested Dependencies in Dutch

3.1 The Fourth Construction

The phrase structure rules discussed so far account for crossed dependencies in Dutch. The literature on dependency relations in Dutch and German (e.g. Bach et al. 1986) has implied that crossed dependency structures are a feature of Standard Dutch, whereas nested dependency structures are a feature of Standard German. However, it appears that the facts are not necessarily this straightforward. Let us now turn to an up-to-now not discussed case of a nested dependency construction in Dutch. The construction I address here is neither a case of verb raising nor of the third construction, and I will refer to it as *the fourth construction* following the tradition that Besten et al. (1988) started. An example of this fourth construction is given in (15). Even though this structure appears to be similar to verb raising constructions, it differs from the verb raising construction in that only a part of the verbal cluster of the complement appears in the position to the right of the matrix verb, instead of the whole verbal cluster.[5] So in (15), the past participle *gevoed* 'fed' is separated from its auxiliary *worden* 'become' (even though they are part of the same complement) by the main predicate *zag* 'saw'. This has consequences for the dependency structure, which as a result is nested: the dependency lines in (15) show the dependency relation between the past participle *gevoed* 'fed' and its argument *een lammetje* 'a lamb', and the dependency relation between the main predicate *zag* 'saw' and its argument *Jan*.

15. *...dat Jan een lammetje gevoed zag worden*
    ...dat Jan a lamb fed saw to-become
    '...that Jan saw a lamb being fed'

I take a non-raising analysis of the passive, following Falk (2003), and I will assume that the auxiliary and the lexical verb correspond to the same minimal f-structure. Under this analysis, the lexical verb is the functional head of the clause and the auxiliaries serve simply as elements that provide voice information. This means that the dependency lines are drawn between the arguments and the lexical verbs and not between the arguments and the auxiliaries.

In other words, the past participle in these constructions does not appear to behave according to the traditional rules of cross-serial dependency constructions. At first sight, the past participle in (15) might appear to be an instance of a cluster creeper, just like other past participles. As said before, past participles are cluster creepers, which are elements in the cluster that creep to various positions within the

[5]This is the case for Standard Dutch, and varieties of Dutch may allow different constructions. One reviewer points out that their southern variety would allow the past particle adjacent to its auxiliary: *...dat Jan een lammetje zag gevoed worden* and would even prefer this order for the verb *laten* 'let': *...dat Jan een lammetje laat gevoed worden*.

cluster, as shown in (14), in which the particle *gezongen* 'sung' can be separated from the verb and appear in two other positions. However, the past participle in the fourth construction cannot be analyzed as a cluster creeper. When comparing a cluster creeper particle to the past participle found in the fourth construction, the difference between the two elements becomes clear: the sentences in (16) show that the past participle *gevoed* (in boldface) can only appear next to its argument and not in any of the other positions that cluster creepers can appear in.

16. (a) *...dat Jan een lammetje* ***gevoed*** *zag worden*
...that Jan a lamb fed saw to-become
'...that Jan saw a lamb being fed'

(b) **...dat Jan een lammetje zag* ***gevoed*** *worden*
...that Jan a lamb saw fed to-become

(c) **...dat Jan een lammetje zag worden* ***gevoed***
...that Jan a lamb saw to-become fed

3.2 Restrictions

There are a few further constraints on this construction. First of all, the fourth construction only occurs in sentences with a perception verb or a causative verb as the matrix verb. Since these verbs are at the same time the only verbs in Dutch that appear with the AcI-construction (Accusativus cum Infinitivo), I refer to them as the AcI-verbs (see *zag* in (17) and (19)). Secondly, fourth the construction will only occur if the XCOMP consists of a passive, which in Dutch is constructed with either the auxiliary *worden* 'become' or *zijn* 'be' in combination with a past participle (see *gevoed worden* 'be fed' in (17) and *geslagen worden* 'be hit' in (18)).

The sentences below show that the matrix verb has to be either a perception verb or a causative verb, and another verb type will be ungrammatical. Only a passive construction (as in (18)) or only a perception verb (as in (19)) is not enough for the fourth construction to be grammatical.

17. Passive, **AcI Verb**

...dat Jan een lammetje gevoed ***zag*** *worden*
...that Jan a lamb fed saw to-become
'...that Jan saw a lamb being fed'

18. Passive, No AcI Verb

**...dat Jan Marie geslagen wilde worden*
...that Jan Marie hit want to-become

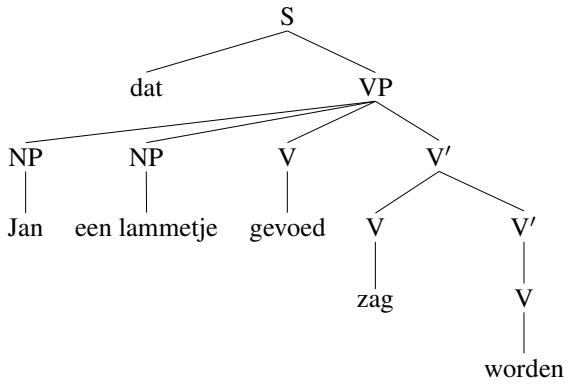
19. No Passive, **AcI Verb**

...dat Jan de vrouw gestoft* ***zag* *hebben*
...that Jan the woman dusted saw to-have

4 Modifications to Kaplan and Zaenen

The c-structure and f-structure that correspond to the fourth construction in (20) are given in Figure 6, in which the passive is analyzed as the functional head of the clause.

20. *...dat Jan een lammetje gevoed zag worden*
...that Jan a lamb fed saw to-become
'...that Jan saw a lamb being fed'



$$\begin{bmatrix} \text{PRED} & \text{'zien } \langle \text{SUBJ, XCOMP} \rangle \text{ OBJ'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'Jan'} \end{bmatrix} \\ \text{TENSE} & \text{PAST} \\ \text{VFORM} & \text{TENSED} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'een lammetje'} \end{bmatrix} \\ \text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{'voeden} \langle \text{SUBJ} \rangle \text{'} \\ \text{SUBJ} & [\ ] \\ \text{VFORM} & \text{PART} \\ \text{AUX-FORM} & \text{WORDEN} \\ \text{PASSIVE} & + \end{bmatrix} \end{bmatrix}$$

Figure 6: C-structure and f-structure for (20)

The c-structure shows the unique aspect of the fourth construction: the separated verbs from the XCOMP verb cluster (i.e. the verb *gevoed* 'fed' and its auxiliary *worden* 'become'), are on either side of the matrix verb *zag* 'saw'. In this analysis, the past participle is placed directly under the VP, in order to ensure this fixed position for the past participle of a passive construction when the main clause's PRED is a perception verb or causative. In order to arrive at this order, the phrase structure rules discussed so far will have to be modified, as shown in Figure 7.

VP	→	NP*	**(V)**	V′	(VP)
		(↑ XCOMP* (COMP) NGF)=↓	(↓ **VFORM)=$_c$PART**		(↑ XCOMP* COMP)=↓
			(↑ XCOMP)=↓		

V′	→	(PostP)	(V)	[V ,	(V′)]
		(↑ XCOMP$^+$ PRT)=↓	(↓ VFORM)=$_c$PART		(↑ XCOMP)=↓
			(↑ XCOMP$^+$)=↓		(↑ XCOMP$^+$ NGF) $\neg <_f$ (↑ NGF)
			↓ $\neg <_f$ (↓ PRT)		
			¬ **(↑ PASSIVE)**		

Figure 7: Modified phrase structure rules for VP and V′

The newly introduced item in the first rules in Figure 7 is the optional V, which is constrained to a past participle that is the head of the XCOMP.

The modification to the second rule in Figure 7 is the constraint $\neg(\uparrow$PASSIVE), which needs to be added to the leftmost V in the V-bar rule, with the assumption that the auxiliary *worden* 'become' in (20) will add [PASSIVE $= +$] to the XCOMP. This constraint is necessary to ensure that the past participle under V-bar cannot be passive. This will correctly disallow the past participle of (20) to appear in the three possible positions of *gezongen* 'sung' as in (14), but permit any other past participle to do so.

The lexical entries for the verbs in (20) are the following:

21. zag
($\uparrow$ PRED) = 'zien <SUBJ, XCOMP> OBJ'
($\uparrow$ TENSE) = PAST
($\uparrow$ VFORM) = TENSED
($\uparrow$ OBJ) = ($\uparrow$ XCOMP SUBJ)
($\uparrow$ XCOMP VFORM) $=_c$ PART
$\uparrow <_f$ ($\uparrow$ XCOMP AUX-FORM)

22. gevoed
($\uparrow$ PRED) = 'voeden <SUBJ>'
($\uparrow$ AUX-FORM) = WORDEN
($\uparrow$ VFORM) = PART

23. worden
($\uparrow$ PASSIVE) = $+$
($\uparrow$ AUX-FORM) = $\downarrow$
$\downarrow$ = WORDEN

The lexical entry for the verb *zag* 'saw' in (21) shows that the verb takes a subject and an XCOMP, and an athematic object, following Kaplan and Zaenen (2003). Its tense is past, its form is tensed, its object is the subject of the XCOMP, emphasizing the object raising nature of these verbs. The form of the predicate of the XCOMP is past participle. The f-precedence rule makes sure that the main predicate will precede the XCOMP's AUX-FORM. The lexical entry for the verb *gevoed* 'fed' in (22) shows that this verb takes a subject. Its auxiliary is *worden* 'become' and its form is that of the past participle. Finally, the lexical entry for the auxiliary *worden* 'become' is given in (23) and contributes the feature PASSIVE to the f-structure of its head. The last two equations are presented in what appears to be a rather indirect manner. However, in order to allow for the f-precedence rule for (21) to work (as f-precedence rules only hold between two f-structures), the value of the AUX-FORM feature needs to be assigned an f-structure. This f-structure corresponds to *worden*.

5 Resultative Adjectives

Interestingly, it appears that this fixed position of the past participle corresponds to the position of resultative adjectives (see (24a)), and resultative PPs (see (24b)).

These resultative adjectives and resultative PPs within an XCOMP appear before the main predicate, whereas the XCOMP's PRED appears after the main predicate. These adjectives cannot appear in any other position. This position appears to be reserved for particles as well (for instance, see Neeleman 1994).[6]

24. (a) *...dat Jan het hek* ***groen*** *wil verven*
...that Jan the fence green wants to-paint
'...that Jan wants to pain the fence green'

(b) *...dat Jan het hout* ***in stukken*** *wil hakken*
...that Jan the wood in pieces want to-chop
'...that Jan wants to chop the wood in pieces'

It therefore appears that the fixed position under VP is not just reserved for past participles of passive constructions, but also for resultatives. This link between past participles and resultatives is reminiscent of Kibort (2005)'s discussion of the *resultative* and *passive* in English. Her term *resultative* refers to a resultative participle (e.g. *broken* in *The vase appeared broken*) that is the complement of the main PRED (e.g *be/seem/appear*), whereas the term *passive* refers to a resultative participle (e.g. *broken* in *The vase got broken*) that fulfills the role of the main PRED itself, taking an auxiliary (e.g. *be/become/get*). These descriptions appear to match those for the past participle form of Dutch passives (e.g. *gevoed*), which contribute their own PRED, on the one hand, and for the Dutch resultative adjectives and resultative PPs, which are XCOMPs of other lexical verbs, following Simpson (2006)'s approach to resultative attributes, on the other.

The Dutch data presented here contribute to this discussion on the blurred distinction between adjectives and participles, as resultative adjectives and past participles appear to share the same syntactic position in Dutch. Whether this shared position follows from some shared property (e.g. the position where predicative elements go) remains an unanswered question and provides a direction for further research.

6 Conclusion

In conclusion, I have shown that the rules by Kaplan and Zaenen (2003) need to be modified to account for the nested dependency structure found in sentences with perception verbs/causatives as the main predicate and a passive construction in the XCOMP. By implementing small changes to the existing phrase structure, the fixed position of the past participle is ensured. The lexical entries in addition will make sure the relative order within the verb cluster is correct and that all the necessary features are assigned to the f-structure.

[6]One reviewer notes that the generalization to resultatives proposed here also holds for their southern variety of Dutch: just as the past participle can appear to the right of the main predicate in this variety, so can the resultative.

I have assumed a non-raising analysis of passives, following Falk (2003), unlike previous work on serial dependency relations, which has always taken a raising approach to modals and auxiliaries.

Finally, I have highlighted a direction for further research, namely the shared syntactic position between the past participle of perception verbs and causative verbs on the one hand, and resultative adjectives and resultative PPs on the other.

References

Bach, Emmon, Colin Brown, and William Marslen-Wilson. 1986. Crossed and Nested Dependencies in German and Dutch: A Psycholinguistic Study. *Language and Cognitive Processes*, 1(4):249–262.

Berman, Judith. 2000. *Topics in the Clausal Syntax of German*. PhD thesis, University of Stuttgart.

Besten, Hans den, Jean Rutten, Tonjes Veenstra, and Joop Veld. 1988. Verb Raising, Extrapositie en de Derde Constructie. Manuscript, UvA.

Bresnan, Joan, Ron M. Kaplan, Stanley Peters, and Annie Zaenen. 1982. Cross-Serial Dependencies in Dutch. *Linguistic Inquiry*, 13:613–36.

Evers, Arnold. 1975. *The Transformational Cycle of Dutch and German*. PhD thesis, University of Utrecht.

Falk, Yehuda. 2003. The English Auxiliary System Revisited. In Butt, Miriam and Tracy Holloway King, (eds.), *Proceedings of the LFG03 Conference*. Stanford: CSLI Publications.

Hinrichs, Erhard and Tsuneko Nakazawa. 1998. Third Construction and VP Extraposition in German. In Erhard Hinrichs, Andreas Kathol and Tsuneko Nakazawa, (eds.), *Complex Predicates in Nonderivational Syntax*. New York: Academic Press.

Kaan, Edith and Nada Vasic. 2004. Cross-Serial Dependencies in Dutch: Testing the Influence of NP-Type on Processing Load. *Memory & cognition*, 32(2): 175–184.

Kaplan, Ron M. and Annie Zaenen. 2003. West-Germanic Verb Clusters in LFG. In Seuren, Pieter and Gerard Kempen, (eds.), *Verb Constructions in German and Dutch. The Series is Current Issues in Linguistic Theory, 242*.

Kibort, Anna. 2005. The Ins and Outs of the Participle-Adjective Conversion Rule. In Butt, Miriam and Tracy Holloway King, (eds.), *Proceedings of the LFG05 Conference*, pages 205–225. Stanford, CA: CSLI Publications.

Neeleman, Ad. 1994. *Complex Predicates*. Utrecht: OTS.

Rambow, Owen. 1992. A Linguistic and Computational Analysis of the German 'Third Construction'. In *Proceedings of the 30th Annual Meeting on Association for Computational Linguistics*.

Rentier, Gerrit. 1994. Dutch Cross Serial Dependencies in HPSG. In *Proceedings of the 15th Conference on Computational Linguistics*, pages 818–822.

Schuurman, Ineke. 1991. Functional Uncertainty and Verb-Raising Dependencies. In Wim Kosmijer, Werner Abraham and Eric Reuland, (eds.), *Trends in Germanic Syntax*. Mouton de Gruyter.

Seuren, Pieter and Gerard Kempen. 2003. *Verb Constructions in German and Dutch*. Amsterdam: Benjamins.

Shieber, Stuart. 1985. Evidence Against the Context-Freeness of Natural Language. *Linguistics and Philosophy*, 8:333–343.

Simpson, Jane. 2006. Resultatives. In Butt, Miriam and Tracy Holloway King, (eds.), *Lexical semantics in LFG*. University of Chicago Press.

Vogel, Carl, Ulrike Hahn, and Holly Branigan. 1996. Cross-Serial Dependencies Are Not Hard to Process. In *Proceedings of the 16th Conference on Computational Linguistics*, pages 157–162.

Zaenen, Annie and Ron Kaplan. 1995. Formal Devices for Linguistic Generalizations: West Germanic Word Order in LFG. In *Formal issues in Lexical-Functional Grammar*, pages 215–239. Stanford: CSLI Publications.

TWO REPRESENTATIONS OF NEGATION IN LFG: EVIDENCE FROM POLISH

Adam Przepiórkowski and Agnieszka Patejuk
Institute of Computer Science, Polish Academy of Sciences

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King (Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

The aim of this paper is to propose f-structure representations of two different types of negation: constituent negation and eventuality negation. In particular, the paper substantiates a suggestion recently made within the PARGRAM community that two binary-valued attributes are needed. We also show that, while the distinction between constituent and eventuality negation is often treated as epiphenomenal in current linguistics, boiling down to scope, Polish provides evidence for the morphosyntactic status of this distinction. Finally, we show that both types of negation may be used metalinguistically, thus supporting Horn's (1985, 1989) analysis of metalinguistic negation as a pragmatic phenomenon.

1 Introduction

There is no standard representation of negation in LFG f-structures; the issue is not mentioned in the two most popular LFG textbooks/reference books, namely Bresnan 2001 and Dalrymple 2001 (or in October 2014 versions of new editions of these monographs). Recently, this topic has raised some interest within the PARGRAM community (`https://pargram.b.uib.no/`), which aims at the development and standardisation of XLE (Crouch et al. 2011) implementations of LFG grammars for various languages. There, two solutions have been proposed: to represent negation as an appropriate element of the ADJ(UNCT) feature, or to introduce a new binary feature, NEG. The following two simplified f-structures illustrate these two possibilities for the sentence "John doesn't like Mary":

(1)
$$\begin{bmatrix} \textsc{pred} & \text{`}\textsc{like}\langle \boxed{1},\boxed{2}\rangle\text{'} \\ \textsc{subj} & \boxed{1}\begin{bmatrix}\textsc{pred} & \text{`}\textsc{John}\text{'}\end{bmatrix} \\ \textsc{obj} & \boxed{2}\begin{bmatrix}\textsc{pred} & \text{`}\textsc{Mary}\text{'}\end{bmatrix} \\ \textsc{adj} & \left\{\begin{bmatrix}\textsc{pred} & \text{`}\textsc{not}\text{'} \\ \textsc{adj-type} & \textsc{neg}\end{bmatrix}\right\} \end{bmatrix}$$

(2)
$$\begin{bmatrix} \textsc{pred} & \text{`}\textsc{like}\langle \boxed{1},\boxed{2}\rangle\text{'} \\ \textsc{subj} & \boxed{1}\begin{bmatrix}\textsc{pred} & \text{`}\textsc{John}\text{'}\end{bmatrix} \\ \textsc{obj} & \boxed{2}\begin{bmatrix}\textsc{pred} & \text{`}\textsc{Mary}\text{'}\end{bmatrix} \\ \textsc{neg} & + \end{bmatrix}$$

[†]We are grateful to both reviewers for their comments, which led to various improvements in the form and content of this paper. Work reported here has been partially financed by the Polish Ministry of Science and Higher Education within the CLARIN ERIC programme 2015–2016 (`http://clarin.eu/`).

Within a particular implementation, usually one of these representations is assumed: in the majority of cases this is the former representation, treating occurrences of negation as adjuncts, because it makes it easy to represent multiple negation (via multiple negative elements of the ADJ set), as in "John doesn't not like Mary". A recent exemplar of this approach is the analysis of Hungarian negation in Laczkó 2014, where all types of negation, including constituent negation and verbal negation, are treated this way. It has also been suggested (Laczkó 2015) that perhaps both kinds of representation are needed in the analysis of Hungarian and, in passing, that perhaps two binary-valued attributes would suffice for this purpose: NEG (with values '+' and '–') and POL (with values such as 'negative' and 'affirmative'). The current proposal may be seen as fleshing out and further substantiating this suggestion.

2 Eventuality negation

For reasons given below, instead of the usual terms *predicate negation* or *sentential negation*, we adopt here the term *eventuality negation* (EN), whose denotation is a little broader than that of these more common terms. In Polish, the usual surface realisation of EN is as the verbal prefix *nie*, e.g.:[1]

(3) Janek nie lubi Marii.
Janek.NOM NEG likes Maria.GEN
'Janek doesn't like Maria.'

(4) Nikt nie lubi nikogo.
nobody.NW.NOM NEG likes nobody.NW.GEN
'Nobody likes anybody.'

Multiple arguments for the bound morpheme status of *nie* – contradicting Polish orthographic rules that treat it as a separate word delimited by spaces – are given in Kupść and Przepiórkowski 2002 and they involve joint prosody, impact on valence frames, paradigm gaps, no scope over coordination (cf. King 1995 for similar considerations in the context of Russian) and the strong adjacency requirement holding between *nie* and the following verb. In fact, Polish orthography rules are a little inconsistent here (some (de)verbal forms are written together with *nie*), they are unstable (the rules about writing *nie* with participles changed a few years ago) and they differ from the – more reasonable, in our opinion – orthography rules for Czech, another West Slavic language, where the grammatical facts discussed below are similar (with the exception of the Genitive of Negation, now extinct in this language), but the Czech negative marker *ne* is consistently written together with the following verbal form.

[1] An attempt is made to follow Leipzig Glossing Rules (`https://www.eva.mpg.de/lingua/resources/glossing-rules.php`) in this paper. Additionally, NW stands for an n-word (see below).

Eventuality negation displays a number of grammatical characteristics, two of which are illustrated by the two sentences above. First of all, as seen in (3), EN triggers the Genitive of Negation (GoN), a phenomenon where a normally accusative argument of a verb occurs in the genitive case when this verb (or a higher verb within an appropriate domain) is negated – see Patejuk and Przepiórkowski 2014a and references therein. Second, as may be seen in (4), Polish is a Negative Concord (NC) language, where n-words such as *nikt* 'nobody.NOM' and *nikogo* 'nobody.ACC/GEN' are licensed by EN – see the rich HPSG literature on Polish NC, including Przepiórkowski and Kupść 1999 and Richter and Sailer 2004, and references therein.

While it is usually only the (pre)verbal negation that is discussed in the context of Negative Concord, there are two other kinds of negative environments that license n-words and, hence, should be included within the denotation of the term *eventuality negation*. One concerns negated adjectives – adjectives such as *niepodobny* 'unsimilar, unlike' (paired with *podobny* 'similar, alike'), may license n-words, as the following attested[2] example – involving the n-word *żadnego* 'none.GEN' – illustrates:

(5) Jest to absolutnie nowy minerał niepodobny do żadnego z dotychczas
is this.NOM absolutely new mineral NEG:similar to none.NW of so-far
nam znanych.
us.DAT known
'This is an absolutely new mineral, not similar to any known to us so far.'

Similarly, de-adjectival negated adverbs may also license n-words:[3]

(6) Smakuje niepodobnie do żadnego innego zboża.
tastes NEG:similarly to none.NW other grain
'It tastes unlike any other grain.'

The other non-verbal environment licensing n-words is the preposition *bez* 'without', as in (7).[4]

(7) Wygram bez żadnych problemów.
win.1.SG.FUT without none.NW.GEN problems.GEN
'I will win without any problems.'

This n-word-licensing environment is discussed in Przepiórkowski and Kupść 1999, where it is suggested that *bez* may express a propositional content meaning roughly 'to not involve'. We adopt this view here and conclude that the discussed environments justify the broad term *eventuality negation*, alluding to the notion of *eventuality* (Bach 1986), which encompasses both *events* and *states* (the latter expressed

[2] `http://www.geekweek.pl/aktualnosci/19103/mineral-niepodobny-do-niczego-co-znamy` (accessed on 25 September 2015). Attested examples are often cited here in a simplified form.

[3] `http://mycuisine.blox.pl/2012/02/Golabki-wegetarianskie-z-amarantusem.html` (accessed on 25 September 2015).

[4] `http://sportowefakty.wp.pl/boks/466592/tomasz-adamek-jesli-bede-szybki-to-wygram-bez-zadnych-problemow` (accessed on 25 September 2015).

not only by verbs, but also by adjectival elements), as opposed to the more common terms *predicate negation*, *sentential negation* or *verbal negation*.[5]

3 Constituent negation

Polish *constituent negation* (CN), while expressed by the same form *nie*, displays markedly different properties than eventuality negation. First, it is not a bound morpheme: it may be separated from the constituent it negates, it may scope over coordination, etc. Second, it does not display the grammatical properties discussed above:

(8) Nie Janek lubi Marię / *Marii / *nikogo (lecz
NEG Janek.NOM likes Maria.ACC Maria.GEN nobody.NW.ACC/GEN but
Tomek).
Tomek.NOM
'It's not Janek who likes Maria (but Tomek).'

As (8) shows, CN does not trigger the Genitive of Negation (*Marię* must occur in the accusative) and it does not license n-words. This is true regardless of the category of the negated constituent; for example, the following attested[6] sentence shows that the otherwise obligatory local GoN does not occur when the form *nie* preceding the verb is interpreted as CN (the clear negative judgement ours):

(9) Ma skakać, a nie pisać wiersze / *wierszy.
has jump.INF and NEG write.INF poems.ACC/*GEN
'He is to jump, and not to write poems.' (of a sportsman)

Similarly, CN does not license n-words even if they occur within the negated constituent, as the following example, constructed on the basis of (10), illustrates:

(10) *Ma skakać, a nie pisać żadne wiersze.
has jump.INF and NEG write.INF none.NW poems.
'He is to jump, and not to any write poems.' (intended)

4 Two f-structure representations of negation

The contrast between (9) above and the attested[7] (slightly modified, with the clear negative judgement ours) sentence (11) below, involving EN, shows that the two kinds of negation should be distinguished at f-structure, if the analysis of case assignment is to be based on f-structure representations (as in Patejuk and Przepiórkowski 2014a):

[5] See also Przepiórkowski 1999 for further defence of the eventuality-level approach to so-called sentential negation.

[6] `http://www.dobramama.pl/pokaz/458957/Piotr_zyla/1/old` (accessed on 25 September 2015).

[7] `http://biuroliterackie.pl/ksiazki/male-cienie-wielkich-czarnoksieznikow-2/opinie/` (accessed on 25 September 2015).

(11) Poetyckim marzeniem Karpowicza było: nie pisać
poetic.INST dream.INST Karpowicz.GEN was NEG write.INF
wierszy / *wiersze.
poems.GEN/*ACC
'The poetic dream of Karpowicz was not to write poems.'

In (11), the nominal predicative phrase occurs preverbally in the instrumental case, and the postverbal subject of the predication is a negated infinitival clause. Just as in (9), the negative marker *nie* directly precedes an infinitival verb, and yet the normally accusative object of this verb must occur in the genitive here.

N-words behave in a similar way, i.e. they are licensed by EN and not by CN, as illustrated by the contrast between (10) above and (12) below.

(12) Poetyckim marzeniem Karpowicza było: nie pisać żadnych wierszy.
poetic.INST dream.INST Karpowicz.GEN was NEG write.INF none.NW poems.
'The poetic dream of Karpowicz was not to write any poems.'

Again, Negative Concord seems to be best handled at the level of f-structure (see the following section), so the two kinds of negation should have different f-structure representations.

The simplest solution consistent with the above facts would consist in positing a single attribute with three possible values corresponding to: no negation, EN and CN. However, CN and EN may co-occur at the same constituent (and, hence, the same f-structure), as the following attested[8] example shows:

(13) Kościół katolicki nie nie potrafi, ale nie chce.
church.NOM catholic.NOM NEG NEG can but NEG want
'It's not that the Catholic Church cannot, but rather that it doesn't want to.'

For this reason we propose to represent negation via two binary attributes: ENEG (for eventuality negation) and CNEG (for constituent negation). Adopting this assumption, the first part of (13), *Kościół katolicki nie nie potrafi* 'It's not that the Catholic Church cannot', will receive the following f-structure:[9]

(14)
$$\begin{bmatrix} \textsc{pred} & \text{'CAN}\langle \boxed{1}, \boxed{2} \rangle\text{'} \\ \textsc{subj} & \boxed{1}\begin{bmatrix} \textsc{pred} & \text{'CC'} \end{bmatrix} \\ \textsc{xcomp} & \boxed{2} \\ \textsc{eneg} & + \\ \textsc{cneg} & + \end{bmatrix}$$

Note that just as we are not aware of the possibility of two eventuality negations occurring at the level of the same predicate, also stacking constituent negation does not seem possible, as the following putative exchange shows:

[8] In the National Corpus of Polish (`http://nkjp.pl/`; Przepiórkowski et al. 2011, 2012); many similar examples may be found in the Internet, e.g. `http://forum.gazeta.pl/forum/w,46,77797868,77846609,nie_nie_potrafi_a_nie_chce_bo_nie_ma_sily_.html?wv.x=1` (accessed on 25 September 2015).

[9] This assumes the ellipsis of XCOMP.

(15) A. Kto lubi Marię?
who.NOM likes Maria.ACC
'Who likes Mary?'

B. Marię lubi Janek.
Maria.ACC likes Janek.NOM
'Janek likes Maria.'

C. Marię lubi nie Janek, lecz Tomek.
Maria.ACC likes NEG Janek.NOM but Tomek.NOM
'It's not Janek who likes Mary, but Tomek.'

B. *Marię lubi nie nie Janek, lecz właśnie on – Janek!
Maria.ACC likes NEG NEG Janek.NOM but FOCUS.PARTICLE him Janek.NOM
'It's not that it's not Janek who likes Mary – he does!' (intended)

The reason the final English translation is grammatical is that negation operates here at two different levels: one negates the whole sentence *that it's…* and the other negates the NP *Janek*. Obviously, this is also marginally possible in Polish, again, with two negation markers *nie* applying to two different constituents:

(16) Nie: Marię lubi nie Janek; Marię lubi właśnie on!
NEG Maria.ACC likes NEG Janek.NOM Maria.ACC likes FOCUS.PARTICLE he.NOM
'It's not that it's not Janek who likes Mary – he does like Mary!'

Hence, representing negation via the two binary attributes, ENEG and CNEG, rather than via an arbitrary number of adjuncts marked as NEG, as in (1), does not only make the difference between eventuality and constituent negation explicit, but also leads to more restrained analyses, where up to one negation of either type is possible for each predicate.

In the context of English, the distinction between constituent negation and sentential (here: eventuality) negation was discussed and supported with various tests in Klima 1964, as in the following minimal pairs (here on the basis of Penka 2015: 304 and Zeijlstra 2015: 275):

(17) **sentential negation:**

John didn't find a job,

a. …did he / *didn't he?
b. …and neither did Mary / *and so did Mary.
c. …and Mary didn't, either / *and Mary didn't, too.
d. …not even a part-time one / *even a part-time one.

(18) **constituent negation:**

John found a job not far away,

a. …didn't he / *did he?

b. …and so did Mary / *and neither did Mary.

c. …and Mary did, too / *and Mary did(n't), either.

d. …even a well-paid one / *not even a well-paid one.

However, since then, the distinction has fallen into disrepute and it has been reanalysed as a simple matter of scope of the negative marker (Payne 1985, Acquaviva 1997; see also Zeijlstra 2015: 275–276 and Penka 2015: 304–305).

This position does not seem to be justified in the case of Polish, where a number of diagnostics distinguish constituent negation from eventuality negation, even when both scope over the predicate. Two such differences, concerning Genitive of Negation and Negative Concord, are illustrated with the minimal pairs (9) vs. (11) and (10) vs. (12). Another difference is the possibility to insert additional material between *nie* and the negated element: as noted in Kupść and Przepiórkowski 2002, it is allowed in CN, but not in EN:

(19) Ma skakać, a nie, kurwa, pisać wiersze!
has jump.INF and NEG whore write.INF poems.ACC/*GEN
'He is to jump, and not to write poems, for fuck's sake!' (of a sportsman)

(20) *Poetyckim marzeniem Karpowicza było: nie, kurwa, pisać wierszy!
poetic.INST dream.INST Karpowicz.GEN was NEG whore write.INF poems.
'The poetic dream of Karpowicz was not to write poems, for fuck's sake!'
(intended)

Also other differences discussed in Kupść and Przepiórkowski 2002 apply here, including different prosodic contours. Such differences between the two kinds of negation in Polish lead Kupść and Przepiórkowski (2002) to the conclusion that the negative marker *nie* in eventuality negation should be treated as a verbal prefix. There is nothing strange about this – the World Atlas of Language Structures (Dryer 2013) reports some 400 languages with negative affixes.[10] Also this is no reason to try to assimilate Polish EN with lexical negation, as expressed in English by the prefixes *un-* (e.g. *unhappy*) and *in-* (e.g. *incontinent*) – such English prefixes have a markedly different semantics than the usual sentential (or eventuality) negation: they express contrary rather than contradictory negation (Horn 1989: §5.1).

In summary, we conclude that the dichotomy discussed here cannot easily be reduced to semantic scope, nor can eventuality negation be equated with so-called lexical negation, and maintain that the two kinds of negation discussed here be represented via two binary-valued attributes, say, ENEG and CNEG.

[10]Strangely enough, while it cites other West Slavic languages (Czech and Sorbian) as having a negative affix, it lists Polish as a negative marker language.

5 A note on Negative Concord

An analysis of case assignment to objects in Polish – including a detailed analysis of the Genitive of Negation – which is compatible with the representation of negation proposed above is presented in Patejuk and Przepiórkowski 2014a, where the attribute NEG is used instead of the current ENEG. Here, we sketch an analysis of Polish Negative Concord. Both analyses are implemented as part of the Polish XLE grammar POLFIE (`http://zil.ipipan.waw.pl/LFG`; Patejuk and Przepiórkowski 2012b,a, 2014b, 2015).

Consider (4), repeated below for convenience, and its intended f-structure in (21).

(4) Nikt nie lubi nikogo.
nobody.NW.NOM NEG likes nobody.NW.GEN
'Nobody likes anybody.'

(21)
$$\begin{bmatrix} \textsc{pred} & \text{'}\textsc{like}\langle\boxed{1},\boxed{2}\rangle\text{'} \\ \textsc{subj} & \boxed{1}\begin{bmatrix} \textsc{pred} & \text{'}\textsc{nobody}\text{'} \\ \textsc{case} & \textsc{nom} \end{bmatrix} \\ \textsc{obj} & \boxed{2}\begin{bmatrix} \textsc{pred} & \text{'}\textsc{nobody}\text{'} \\ \textsc{case} & \textsc{gen} \end{bmatrix} \\ \textsc{eneg} & + \end{bmatrix}$$

The following (relevant parts of) two lexical entries are assumed for the negative marker *nie*, where both features ENEG and CNEG are assumed to be instantiated (may be assigned a value only once):

(22) *nie* ENEG $(\uparrow \textsc{eneg}) = +$

(23) *nie* CNEG $(\uparrow \textsc{cneg}) = +$

Two different preterminal categories ENEG and CNEG are needed, as the two kinds of negative marker have different distribution and occur in different c-structure rules; simplified examples of such rules are given below:

(24) IP $\rightarrow$ (CNEG) (ENEG) I

(25) XP $\rightarrow$ CNEG XP

We adopt here the convention that nonterminals on the right-hand side which are not explicitly annotated with functional equations are implicitly annotated with the head equation $\uparrow{=}\downarrow$; in effect, all nonterminals on the right-hand sides of the above rules are co-heads. The first of these rules says that a (widely understood) verbal category may be preceded by constituent negation and by eventuality negation. If both occur, as in (13) above, the f-structure corresponding to IP (and to I) will include the + values of both CNEG and ENEG. If *nie* occurs only once, as in (4), it may be interpreted as either constituent negation or eventuality negation, so ambiguity arises. In the actual implementation, this ambiguity is handled by adding

OT marks[11] to the effect that eventuality interpretation is preferred. Conversely, the second of the two rules says that CNEG – and only CNEG – may precede a number of maximal projections – XP stands for NP, PP, AP, AdvP, etc. Due to the fact that CNEG is an instantiated feature, this rule may be used only once per maximal projection.

In case of n-words such as *nikt* 'nobody.NOM' and *nikogo* 'nobody.ACC/GEN', the following lexical entries are assumed:[12]

(26) *nikt* N (↑ CASE) = NOM
((XCOMP* GF+ ↑) ENEG) =$_c$ +

(27) *nikogo* N (↑ CASE) ∈ {ACC, GEN}
((XCOMP* GF+ ↑) ENEG) =$_c$ +

The following (here simplified) definition of GF is assumed here; the complete definition also includes various subtypes of OBL and OBJ, but – crucially – it does not include COMP or XCOMP:

(28) GF ≡ {SUBJ|OBL|OBJ|ADJ ∈}

In case of example (4) and its f-structure (21), the functional uncertainty XCOMP* GF+ matches SUBJ (for *nikt*) and OBJ (for *nikogo*). The complexity of the inside-out functional uncertainty present in these lexical entries arises from examples such as the following, where the path specification matches XCOMP OBJ ADJ ∈.

(29) Karpowicz nie chciał pisać żadnych wierszy.
Karowicz.NOM NEG wanted write.INF none.NW poems.GEN
'Karpowicz didn't want to write any poems.'

See Przepiórkowski and Kupść 1997 for a further discussion of locality constraints on the Negative Concord in Polish.

6 A note on metalinguistic negation

Apart from sentential, constituent and lexical negation (the last one not analysed here), there is another type of negation widely recognised in the linguistic literature, namely, metalinguistic negation (Horn 1985, 1989), illustrated here with the following examples from Horn 1989: ch.6:

(30) The king of France is not bald—(because) there is no king of France.

(31) Some men aren't chauvinists—all men are chauvinists.

(32) He didn't call the [pólis], he called the [polís].

[11] See the XLE documentation at `http://www2.parc.com/isl/groups/nltt/xle/doc/xle.html#Local_Optimality_Marks`.

[12] In the current XLE implementation, which uses so-called sublexical rules (Krasnowska-Kieraś and Patejuk 2015), lexical entries define lexemes rather than particular forms, so (26)–(27) are conflated into a single lexical entry in the implementation, with case information coming from the morphological component.

(33) I didn't manage to trap two mongeese—I managed to trap two mongooses.

(34) We didn't have intercourse—we fucked.

(35) Ben Ward is not a black Police Commissioner but a Police Commissioner who is black.

As these examples show, the impact of metalinguistic negation is not the truth-conditional contradiction of the state of affairs expressed by the affirmative counterpart, but rather raising objection to some aspect of an utterance – in fact, any aspect: classical presupposition, scalar implicature, pronunciation, morphology, register or general connotation.

Horn (1985, 1989) notes that, across languages, such metalinguistic uses of negation do not require the use of a negative morpheme different from that used for ordinary "descriptive" (to use Horn's term) negation. Moreover, as exemplified above, typical uses of such metalinguistic negation are not truth-conditional. For this reason, Horn analyses the descriptive vs. metalinguistic ambiguity of negative markers as pragmatic. Nevertheless, there is an aspect of metalinguistic negation that resembles constituent negation: just as CN does not license n-words in Polish, metalinguistic negation does not license Negative Polarity Items (NPIs) in English – and also does not forbid Positive Polarity Items (PPIs) (Karttunen and Peters 1979: 46–47, Horn 1989: 368):

(36) Chris didn't manage to solve {some/*any} problems—he solved them easily.

(37) Bill hasn't already forgotten that today is Friday, because today is Thursday.

In (36), where what is negated is the implicature introduced by *manage*, namely, that the activity expressed by the following infinitival phrase is difficult, the NPI *any* is not allowed, and instead *some* is used. This should be contrasted with the ordinary negation in *Chris didn't solve any problems*, where *any* is allowed (and *some* would result in wide scope over negation). Similarly, in (37), where what is negated is the implicature concerning the factivity of the sentential complement of *forgotten*, the PPI *already* is allowed, although it is normally forbidden in negated sentences (and *yet* should be used instead), e.g. *Bill hasn't forgotten about that yet/*already*.

Given this similarity between Polish CN, which does not license n-words, and English metalinguistic negation, which does not count as negation for the purpose of NPI/PPI-licensing, it is tempting to say that perhaps Polish constituent negation is a subtype of metalinguistic negation. However, the following (constructed but not controversial) data force us to reject this attempt at applying Occam's razor:

(38) a. Nie odwiedziłem Marysi, tylko Anię.
NEG visited.1.SG.M Marysia.GEN only Ania.ACC
'I didn't visit Marysia, but Ania.'

b. Odwiedziłem nie Marysię, tylko Anię.
visited.1.SG.M NEG Marysia.ACC only Ania.ACC
'I visited not Marysia, but Ania.'

(39) a. Nie zwiedzałem żadnego Tübingen, tylko Tybingę.
NEG sightsee1.SG.M none.NW.GEN Tübingen.GEN only Tybinga.ACC
'I wasn't sightseeing Tübingen, but Tybinga.'

b. Zwiedzałem nie jakieś Tübingen, tylko Tybingę.
sightsee1.SG.M NEG some.ACC Tübingen.ACC only Tybinga.ACC
'I was sightseeing not Tübingen, but Tybinga.'

(40) a. Ta kawa nie była gorąca, tylko wrząca!
this.NOM coffee.NOM NEG was hot.NOM only scalding.NOM
'This coffee wasn't hot – it was scalding hot!'

b. Ta kawa była nie gorąca, tylko wrząca!
this.NOM coffee.NOM was NEG hot.NOM only scalding.NOM
'This coffee was not hot but scalding hot!'

The first of these pairs shows that both EN (in (38a)) and CN (in (38b)) may express the usual truth-conditional meaning. The other two pairs show, on the other hand, that both EN and CN may be used metalinguistically: what is objected in (39) is the use of the international name *Tübingen* instead of its Polish version *Tybinga*, and what is negated in (40) is the scalar implicature (that the coffee was cooler than 'hot') – a typical use of metalinguistic negation.

Note in particular that the metalinguistic negation in (39a) licenses the n-word *żadnego*. This may seem surprising at first, given NPIs and PPIs' indifference to metalinguistic negation, but in fact this only confirms the observation that NPI/PPI-licensing on one hand and Negative Concord on the other operate at different linguistic levels: in NC, *licensing is not merely a question of semantic compatibility with a certain context [as it is in the case of NPI/PPI-licensing], but rather, perhaps primarily, a question of syntax* (Giannakidou 2011: 1684). In other words, metalinguistic negation cancels the usual impact of negation on NPI/PPI-licensing at the pragmatic level, as this level is arguably relevant to NPI/PPI-licensing (see e.g. Giannakidou 2011 and references therein), but it cannot cancel the impact of negation on Polish Negative Concord, as NC is a phenomenon at the level of morphosyntax, syntax and semantics, but not pragmatics. In any case, whether this intuition is valid or not, we may safely conclude on the basis of examples such as (38)–(40) that the distinction between constituent and eventuality negation in Polish is orthogonal to the possibility of using either of them metalinguistically. Since metalinguistic negation is a pragmatic phenomenon (Horn 1985, 1989), we assume that it is not explicitly represented within f-structures, but rather that positive values of f-structure attributes ENEG and CNEG may be interpreted as metalinguistic negation at more pragmatic levels of representation.

7 Conclusion

Citing Polish facts of the kind apparently not discussed in the LFG literature so far, we have shown that two different f-structure representations are needed to success-

fully handle the different behaviour of eventuality negation and constituent negation. We have also argued that these two kinds of negation may both be used metalinguistically. The two representations have been put to test in a comprehensive implemented grammar of Polish, which also includes an exhaustive treatment of case assignment and Negative Concord.

References

Acquaviva, Paolo. 1997. *The Logical Form of Negation: A Study of Operator-Variable Structures in Syntax*. New York: Garland.

Bach, Emmon. 1986. The Algebra of Events. *Linguistics and Philosophy* 9, 5–16.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Blackwell Textbooks in Linguistics, Malden, MA: Blackwell.

Butt, Miriam and King, Tracy Holloway (eds.). 2014. *The Proceedings of the LFG'14 Conference*, Stanford, CA, CSLI Publications.

Crouch, Dick, Dalrymple, Mary, Kaplan, Ron, King, Tracy, Maxwell, John and Newman, Paula. 2011. XLE Documentation. `http://www2.parc.com/isl/groups/nltt/xle/doc/xle_toc.html`.

Dalrymple, Mary. 2001. *Lexical Functional Grammar*. San Diego, CA: Academic Press.

Dryer, Matthew S. 2013. Negative Morphemes. In Matthew S. Dryer and Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*, Leipzig: Max Planck Institute for Evolutionary Anthropology.

Giannakidou, Anastasia. 2011. Negative and Positive Polarity Items. In Claudia Maienborn, Klaus von Heusinger and Paul Portner (eds.), *Semantics: An International Handbook of Natural Language Meaning*, volume 2, pages 1660–1712, Berlin: De Gruyter Mouton.

Horn, Laurence R. 1985. Metalinguistic Negation and Pragmatic Ambiguity. *Language* 61(1), 121–174.

Horn, Laurence R. 1989. *A Natural History of Negation*. Chicago, IL: Chicago University Press.

Karttunen, Lauri and Peters, Stanley. 1979. Conventional Implicature. In Choon-Kyu Oh and David A. Dinneen (eds.), *Presupposition*, volume 11 of *Syntax and Semantics*, pages 1–56, Bloomington, IN: Academic Press.

King, Tracy Holloway. 1995. *Configuring Topic and Focus in Russian*. Stanford, CA: CSLI Publications.

Klima, Edward. 1964. Negation in English. In Jerry A. Fordor and Jerrold Katz (eds.), *The Structure of Language*, pages 246–323, New Jersey: Prentice Hall.

Krasnowska-Kieraś, Katarzyna and Patejuk, Agnieszka. 2015. Integrating Polish LFG with External Morphology. In Markus Dickinson, Erhard Hinrichs, Ag-

nieszka Patejuk and Adam Przepiórkowski (eds.), *Proceedings of the Fourteenth International Workshop on Treebanks and Linguistic Theories (TLT 14)*, pages 134–147, Warsaw: Institute of Computer Science, Polish Academy of Sciences.

Kupść, Anna and Przepiórkowski, Adam. 2002. Morphological Aspects of Verbal Negation in Polish. In Peter Kosta and Jens Frasek (eds.), *Current Approaches to Formal Slavic Linguistics: Proceedings of the Second European Conference on Formal Description of Slavic Languages, Potsdam, 1997*, pages 337–346, Frankfurt am Main: Peter Lang.

Laczkó, Tibor. 2014. Outlines of an LFG-XLE Account of Negation in Hungarian Sentences. In Butt and King (2014), pages 304–324.

Laczkó, Tibor. 2015. On an LFG-XLE Treatment of Negation in Hungarian, slides presented at the PARGRAM meeting in Warsaw in February 2015; `http://typo.uni-konstanz.de/redmine/attachments/download/30/laczko_negation_ParGram_Warsaw_140204.pdf`.

Patejuk, Agnieszka and Przepiórkowski, Adam. 2012a. A Comprehensive Analysis of Constituent Coordination for Grammar Engineering. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)*, pages 2191–2207, Mumbai, India.

Patejuk, Agnieszka and Przepiórkowski, Adam. 2012b. Towards an LFG Parser for Polish: An Exercise in Parasitic Grammar Development. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2012*, pages 3849–3852, ELRA, Istanbul, Turkey.

Patejuk, Agnieszka and Przepiórkowski, Adam. 2014a. Structural Case Assignment to Objects in Polish. In Butt and King (2014), pages 429–447.

Patejuk, Agnieszka and Przepiórkowski, Adam. 2014b. Synergistic Development of Grammatical Resources: A Valence Dictionary, an LFG Grammar, and an LFG Structure Bank for Polish. In Verena Henrich, Erhard Hinrichs, Daniël de Kok, Petya Osenova and Adam Przepiórkowski (eds.), *Proceedings of the Thirteenth International Workshop on Treebanks and Linguistic Theories (TLT 13)*, pages 113–126, Tübingen, Germany: Department of Linguistics (SfS), University of Tübingen.

Patejuk, Agnieszka and Przepiórkowski, Adam. 2015. Parallel Development of Linguistic Resources: Towards a Structure Bank of Polish. *Prace Filologiczne* LXV, 255–270.

Payne, John. 1985. Negation. In Timothy Shopen (ed.), *Language Typology and Syntactic Description*, pages 197–242, Cambridge: Cambridge University Press.

Penka, Doris. 2015. Negation and Polarity. In Nick Riemer (ed.), *The Routledge Handbook of Semantics*, pages 303–319, London: Routledge.

Przepiórkowski, Adam. 1999. On Negative Eventualities, Negative Concord, and Negative *yes/no* Questions. In Tanya Matthews and Devon Strolovitch (eds.), *Proceeding of Semantics and Linguistic Theory 9*, pages 237–254, Ithaca, NY: CLC

Publications.

Przepiórkowski, Adam and Kupść, Anna. 1997. Unbounded Negative Concord in Polish: A Lexicalist HPSG Approach. In Jan Landsbergen, Jan Odijk, Kees van Deemter and Gert Veldhuijzen van Zanten (eds.), *Computational Linguistics in the Netherlands 1996: Papers from the Seventh CLIN Meeting*, pages 129–143, IPO, Center for Research on User-System Interaction, Eindhoven: Technische Universiteit Eindhoven.

Przepiórkowski, Adam and Kupść, Anna. 1999. Eventuality Negation and Negative Concord in Polish and Italian. In Robert D. Borsley and Adam Przepiórkowski (eds.), *Slavic in Head-Driven Phrase Structure Grammar*, pages 211–246, Stanford, CA: CSLI Publications.

Przepiórkowski, Adam, Bańko, Mirosław, Górski, Rafał L., Lewandowska-Tomaszczyk, Barbara, Łaziński, Marek and Pęzik, Piotr. 2011. National Corpus of Polish. In Zygmunt Vetulani (ed.), *Proceedings of the 5th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, pages 259–263, Poznań, Poland.

Przepiórkowski, Adam, Bańko, Mirosław, Górski, Rafał L. and Lewandowska-Tomaszczyk, Barbara (eds.). 2012. *Narodowy Korpus Języka Polskiego*. Warsaw: Wydawnictwo Naukowe PWN.

Richter, Frank and Sailer, Manfred. 2004. Polish Negation and Lexical Resource Semantics. *Electronic Notes in Theoretical Computer Science* 53, 309–321.

Zeijlstra, Hedde. 2015. The Morpho-Syntactic Realisation of Negation. In Tibor Kiss and Artemis Alexiadou (eds.), *Syntax – Theory and Analysis. International Handbook. Volume 1*, pages 274–309, Berlin: De Gruyter Mouton.

CLAUSAL SUBJECTS AND EXTRAPOSITION IN THE HISTORY OF ENGLISH

Rickard Ramhöj

University of Gothenburg

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King (Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

Using Kibort's revised Lexical Mapping Theory (Kibort, 2007, 2008, 2013, 2014), this paper gives an analysis of the argument structure of verbal predicates that alternately take (i) a clausal subject and (ii) a subject *it* in conjunction with propositional subclause. Based on data from syntactically annotated corpora of historical English, two separate argument structures are posited for the relevant predicates. A distinction is made between, on the one hand, a thematic subject *it* occurring together with an adjunct subclause and, on the other hand, a non-thematic subject *it* occurring together with a complement subclause. This distinction provides an explanation for a number of facts in present-day and historical English concerning extraction and co-occurrence patterns with respect to the predicates showing the alternation under discussion.

1 Introduction

In all periods of the history of English, there have been predicates that alternately occur with, on the one hand, a clausal subject[1] and, on the other hand, a subject *it* in conjunction with a propositional subclause. This alternation, which often goes under the label *it*-extraposition, is exemplified in (1). Both sentences are taken from the Penn-Helsinki Parsed Corpus of Early Modern English (Kroch et al., 2005). The subclause is marked out by square brackets, something that is done throughout the article. Note that the clausal subject is preposed in (1-a). As will be seen in section 4 of the article, in early English, there are also clausal subjects that are not preposed.

(1) a. Preposed clausal subject:
but [that it should go amongst the Rebels] is a strange thing.
(DRUMMOND-E3-P1,2.4,189.10)

b. Subject *it* in conjunction with a subclause:
It is a strange thing [you will impute that to me], ...
(RALEIGH-E2-H,I,209.C1.155)

In (1-a), the *that*-clause *that it should go amongst the Rebels* occurs in a clause-initial position followed by the copula *be* and the nominal phrase *a strange thing*. The construction exemplified by the sentence in (1-a) will here be called the preposed clausal subject construction; *preposed*, because the subclause is clause-initial while subclauses typically occur in a clause-final position; *clausal subject*,

[†]I would like to thank the reviewers for very useful feedback on an earlier version of this article. I would also like to thank the participants at LFG15 in Tokyo, Japan, as well as Gerlof Bouma, Maia Andréasson, Gunnar Bergh, Henrik Rosenkvist, Hubert Cuyckens, Filippa Lindahl, Elisabet Engdahl, Sölve Ohlander, Heike Havermeier and Michelle Waldispühl for valuable comments and suggestions.

[1]The existence of clausal subjects in Old English is a contentious issue. For a study that argues for the nonexistence of clausal subjects in Old English, see Anderson (1997).

because there is evidence to suggest that the subclause in this kind of construction is to be analyzed as a functional subject (Huddleston and Pullum, 2002, 957).

In (1-b), we see the same predicate *be a strange thing* in combination with a subject *it* and the subclause *that you will impute that to me* in clause-final position. The construction exemplified by the sentence in (1-b) will descriptively be referred to as the *it+subclause* construction.

In this paper, it is argued that not all cases of a subject *it* in conjunction with a propositional subclause can be given the same analysis. The aim of the paper is to argue that the *it+subclause* construction can be divided into two different constructions: (i) the *it*+ADJ construction and (ii) the *it*+COMP construction. Consider the difference between the sentences in (1) and the sentences in (2). The sentences in (2) derive from Seppänen and Herriman (2002).

(2) a. *[That there was no alternative] seems.
b. It seems [that there was no alternative].
c. [That we were to take care of the remaining work] seemed like a good idea at the time.

In (1), with the nominal predicate *be a strange thing*, there is an alternation between the preposed clausal subject in (1-a) and the *it+subclause* construction in (1-b). In (2), with the raising verb *seem*, the alternative with a preposed clausal subject is only possible in conjunction with a secondary predicate, as in (2-c), and not on its own, as in (2-a) (Seppänen and Herriman, 2002). Considering this difference between the sentences in (1) and (2), it does not seem possible to give the same analysis to both predicates.

The analysis proposed in this paper is that the sentence in (1-b) represents the *it*+ADJ construction, while the sentence in (2-b) represents the *it*+COMP construction. The *it*+ADJ construction consists of a thematic subject *it* in conjunction with an adjunct subclause, while the *it*+COMP construction consists of a non-thematic subject *it* in conjunction with a complement subclause. Based on the revised Lexical Mapping Theory of Kibort (2007, 2008, 2013, 2014), an analysis is given of the differences between the predicates that occur in these two constructions in present-day and historical English.

In the next section, a short background is given on previous analyses of a subject *it* in conjunction with a subclause in LFG. This is followed by a presentation of my own analysis in section 3. In section 4, the question is discussed whether the argument structures, presented for data on Present-Day English, also fit the data from early English.

2 Previous LFG analyses

Subject *it* in conjunction with a subclause in LFG has been discussed mainly in connection with the phenomenon of raising. In the Lexical Mapping Theory presented in Bresnan et al. (2016), the argument structure and argument-to-function

mapping of the raising verb *seem* is represented, see (3) from (Bresnan et al., 2016, 340).

(3) Argument structure for *seem*:

seem	_	⟨*experiencer*,	*proposition*⟩
	[–r]	[–o]	[–o]
	|	|	|
	SUBJ	OBL_{exp}	XCOMP/COMP

In (3), the verb *seem* takes three arguments, where the first argument constitutes an empty argument role that does not have any semantic content. Following the mapping principles, the empty argument role is mapped to SUBJ, the *experiencer* to OBL_{exp} and the *proposition* to XCOMP or COMP. This argument structure accounts, among other things, for the fact that the subject of the verb *seem* can be non-thematic *there* or *it*. An example of the verb *seem* taking non-thematic subject *there* or *it* is given in (4).

(4) a. There seems to me [to be a problem with the proposal]. (Bresnan et al., 2016, 340)
b. It seems to me [that there is a problem with the proposal]. [constructed]

In (4-a), *there* is the subject of *seem* and also the subject within the XCOMP *to be a problem with the proposal*. The prepositional phrase *to me* is linked to OBL_{exp}. In (4-b), the nonthematic *it* is linked to SUBJ, *to me* to OBL_{exp}, and the subclause *that there is a problem with the proposal* to COMP. The empty subject argument can thus either be filled with a raised argument from an XCOMP, as in (4-a), or with the nonthematic *it*.

The postulation of an empty argument structure slot, which is not associated with a thematic role, implies that there is a distinction to be made between thematic roles and argument structure slots. In the revised Lexical Mapping Theory of Kibort (2007, 2008, 2013, 2014), this distinction between thematic roles and argument structure slots is made explicit in the sense that an independent set of argument structure slots[2] is introduced, separate from the level of thematic roles. Each argument slot is associated with a set of proto-role entailments, governing the possible mappings between thematic roles and argument structure slots (Kibort, 2007, 257). The argument structure of the verb *seem* according to the revised Lexical Mapping Theory is given in (5).

(5) Argument structure for *seem*:

[2]List of argument structure slots:

arg1	arg2	arg3	arg4	… arg_n
[–o]/[–r]	[–r]	[+o]	[–o]	[–o]

	∅	x	y
	|	|	|
seem	arg1	arg4	arg5
	[–r]	[–o]	[–o]
	|	|	|
	SUBJ	OBL_{exp}	XCOMP/COMP

In (5), the verb *seem* takes the three argument slots, arg1, arg4 and arg5, which each is associated with an intrinsic feature, [–r], [–o], and [–o]. According to the mapping principle of the revised LMT[3] (Kibort, 2007), arg1[–r] is mapped to SUBJ, and arg4[–o] and arg5[–o] to obliques (OBL_{exp} and XCOMP[4]). The result of the mapping-to-functions is the same as in (3), but the distinction between thematic roles and argument structure slots gives a more intuitive depiction[5] of the fact that the verb *seem* in sentences such as those in (4) takes three syntactic arguments, but only two thematic roles.

The analyses given above with respect to the raising verb *seem* represent the way in which all instances of the *it+subclause* construction in English have typically been analysed. There is no distinction made between thematic and non-thematic *it* in these constructions. One example of a study that doesn't make this distinction is Haugland (2006, 39), who, when presenting the sentence in (6), holds that the subject *it* is non-thematic and the subclause a complement.

(6) It is true [that John saved her].

Haugland (2006, 39) describes the subject *it* in (6) as a 'semantically empty (or nonreferential) entity used solely for some grammatical, or rather, syntactic purpose'.

One study that does make a distinction between different kinds of subject pronouns in conjunction with propositional subclauses, but in relation to German data, is Berman (2003). Berman (2003) gives two separate analyses for the subject *es* in the sentences in (7).

(7) a. weil es gesagt wurde, [dass Hans krank ist].
because it said was that Hans sick is
'because it was said that Hans is sick.'

b. weil es mich stört, [dass sie den Hans liebt].
because it bothers me that she Hans loves

[3]The mapping principle of the revised LMT is that 'the ordered arguments are mapped on to the highest (i.e. least marked) compatible function on the markedness hierarchy' (Kibort, 2007).

[4]The functions COMP and XCOMP are being equated with OBL_{prop} (cf. Zaenen and Engdahl, 1994).

[5]Kibort's revised Lexical Mapping Theory has additional advantages. Given the proto-role entailments of the argument slots, a certain thematic role could be mapped to different slots, resulting in different semantic entailments. Kibort (2007) gives the example of the *agent* role of a certain Polish predicate, which, as a result of a morphosemantic operation, is mapped to a secondary object, giving rise to an interpretation in which the *agent* is 'unwilful'.

'because it bothers me that she loves Hans.'

Berman (2003) argues that the the sentence in (7-a) has a thematic subject *es* in conjunction with an adjunct subclause, i.e. *it*+ADJ, while the sentence in (7-b) has a non-thematic subject *es* in conjunction with a complement subclause, i.e. *it*+COMP. The evidence given for this distinction comes among other things from *wh*-extraction. The analysis provides an explanation why it is not possible to extract out of the subclause in (8-a), while, according to Berman (2003, 152), extraction from the subclause in (8-b) is perfectly acceptable.

(8) a. *Was wurde es gesagt, [dass er gelesen hat].
What was it said that he read has
'What was it said that he has read.'
b. Wen stört es dich, [dass sie liebt].
Who bothers it you that she loves
'Who does it bother you that she loves.'

In (8-a), passive *gesagt werden* takes a theme subject, which corresponds to the pronoun *es*. The assumption that the predicate here takes only one argument results in an analysis of the subclause as an adjunct. As is standardly assumed, adjuncts constitute syntactic islands out of which extraction is not possible, or at least is significantly more difficult than for complements (Bresnan et al., 2016, 287).

In (8-b), according to Berman (2003, 161), the verb *stören* takes a non-thematic subject *es*, an experiencer object and a theme complement. The *it*+COMP analysis accounts for the grammaticality of extraction out of the subclause, shown in (8-b). Because of the fact that *stören* does not always take a non-thematic subject, the non-thematic subject is treated as optional (Berman, 2003, 163).

In the present paper, the distinction between *it*+ADJ and *it*+COMP, which is discussed in Berman (2003), is applied to Present-Day and Historical English. Some aspects of Berman's analysis are adopted, while others are rejected. A problem with Berman's analysis is the fact that the presence of non-thematic *es* in Berman (2003) seems to a large extent to be the result of lexical idosyncracies. How is it that a verb such as *stören* requires a non-thematic subject while passive constructions don't? How is it that passive constructions sometimes take a covert non-referential subject and sometimes a thematic *es*? In the next section, my analysis for constructions containing a subject *it* in conjunction with a subclause in historical English is presented. As will be seen, this analysis also gives a different account of the German data presented by Berman (2003), where the presence or absence of a non-referential subject is not taken to be the result of lexical idiosyncrasies.

3 The *it*+ADJ and *it*+COMP constructions

The argument structure analysis of the present paper is performed using the revised Lexical Mapping Theory of Kibort (2007, 2008, 2013, 2014). One important aspect

of this theory is that there is no Subject Condition, i.e. not all verbal predicates need subcategorize for SUBJ. The absence of the Subject Condition has consequences for the analysis of non-thematic subjects. Why do certain predicates take a non-thematic subject, when there is no principle that forces subjects to be present? As there is no thematic role associated with the subject, there is no semantic reason for a non-thematic subject to be present. There is also no structural reason that could be posited for the presence of a non-thematic subject. A constraint that a particular phrase-structural position, such as SpecIP, needs to be filled could be formulated solely with respect to the c-structure and needs not involve the requirement for verbs to subcategorize for a subject. In the current section, the *it*+COMP and the *it*+ADJ constructions are discussed in one subsection each.

3.1 The *it*+COMP construction

In the analysis given here, non-thematic subject *it* in English occurs exclusively in connection with raising verbs, including the copula *be* in passive constructions. My analysis of the argument structure of raising verbs is given in (9). This analysis applies both to lexical raising verbs, such as *seem* and *appear*, and to the copula *be* in passive constructions.

(9) The argument structure of raising verbs:

	∅	proposition
	|	|
raising verb	⟨arg1	arg4⟩
	[–r]	[–o]
	|	|
	SUBJ	COMP/XCOMP

In (9), an abstract representation of the argument structure of raising verbs is given. Individual raising verbs might diverge slightly from this representation. As could be seen in (5), the verb *seem* takes an optional oblique experiencer. The characteristic feature of raising verb is that the arg1[– r] slot is not associated with a thematic role. This is true both for lexical raising verbs such as *seem* and for the copula *be* used in passive constructions.

Two pieces of support for the analysis of passive *be* as a raising verb are given here[6]. The first piece of evidence concerns the thematic relation between the main clause subject and an infinitival phrase in a sentence such as (10).

(10) She is said [to have bine the death of her husband]. (MONTAGUE-E3-P2,1,219.78)
'She is said to have been the death of her husband.'

[6]For additional support for the raising analysis of passivization, based on Polish data, see Patejuk and Przepiórkowski (2014).

In the sentence in (10), the constituent *she* appears to be a thematic argument of the predicate *be the death of her husband*, and not a thematic argument of the predicate *be said*. It is not *she* that is said, but rather the proposition that *she is the death of her husband*. My analysis of the argument structure of the passive participle *said* is given in (11).

(11) The argument structure of the passive participle *said*:

	agent	proposition
	|	|
$said_{pass.part}$	⟨arg1	arg4⟩
	[–o]	[–o]
	[+r]	[+r]
	|	|
	(OBL$_{agent}$)	XCOMP/COMP

In the analysis of passive participles, the arg1[–o] is assigned a [+r] feature[7], demoting arg1[–o] to the function of OBL. For the passive participles that take a propositional subclause as an argument, I also assume that the second argument, arg4[–o], is assigned a [+r] feature. This operation demotes the arg4[–o], which would have become SUBJ, to the function XCOMP. The passive participle in (11) is thus subject to two morphosyntactic operations, where a [–o] argument is assigned a [+r] feature. For these passive participles, they thus take two types of oblique arguments, an OBL$_{agent}$ and a XCOMP or COMP, and take no SUBJ[8]. For the purpose of illustration, a simplified f-structure associated with the sentence in (10) is given in Figure 1.

The f-structure in Figure 1 shows a functional equation relationship between the subject of the main clause *be* and the subject of the XCOMP *to have bine the death of her husband*. The functional equation required for this to work differs from the typical functional equation for subject raising (Bresnan et al., 2016, 304) in the sense that the subject that is 'raised' occurs as the subject within an XCOMP which occurs within another XCOMP. Raising *be* in passive constructions thus needs to have the specification in (12).

(12) (↑ SUBJ) = (↑ XCOMP XCOMP SUBJ)

[7]Kibort (2007) assumes three morphosyntactic operations, shown in (i)

(i) Morphosyntactic operations:
   a. adding the [+r] specification to a [–o] argument;
   b. adding the [+o] specification to a [–r] argument; and
   c. adding the [+r] specification to a [+o] argument.

[8]Kibort's revised LMT does not include the Subject Condition, i.e. it does not include the rule that all verbal predicates take a SUBJ. This means that subjectless predicates, such as the passive participle in (11), are allowed. To get the obligatoriness of subjects in English, a c-structure rule must be posited, which says that the SpecIP position must be filled by a SUBJ. Such a rule is required for Present-day English, regardless if the Subject Condition is assumed.

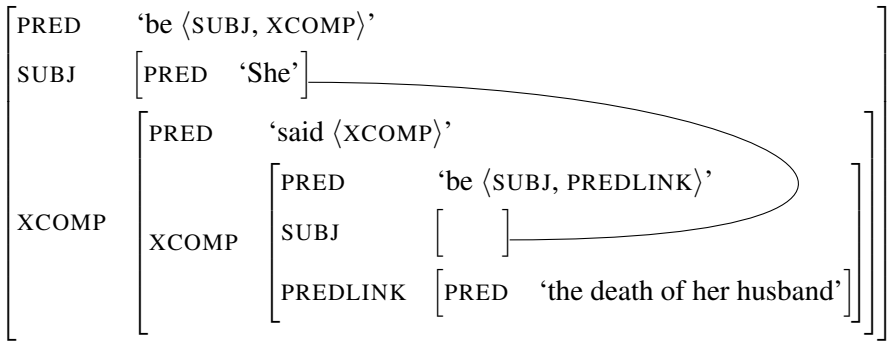


Figure 1: F-structure for the sentence *She is said to have bine the death of her husband*

The equation shown in (12) needs to be present as an alternative to the ordinary raising equation, (↑ SUBJ) = (↑ XCOMP SUBJ), in the lexical entry of raising *be*. The outside-in equation in (12) says that the SUBJ of a clause also constitutes the SUBJ embedded within two XCOMPs.

With the standard analysis of a passive construction, such as *be said*, in which the copula *be* does not have its own PRED value (e.g. Bresnan et al., 2016, 35), an additional [–r] argument needs to be made available when the predicate is in the passive voice, an argument which is not available in the active voice[9].

A second piece of support for the analysis of *be* in passive constructions as a raising verb is that the subject of *be*, just like the raising verb *seem* in (4), allows the two non-thematic subjects *it* and *there*, exemplified in (13).

(13) a. It is said [that Dunkirk is sold to the French for four hundred thousand pound]. (HOXINDEN-1660-E3-H,280.184)
b. there is said [to be in it of Churches & Chappels, 150]. (JOTAYLOR-E2-P2,3,96.C2.299)
'There is said to be 150 churches and chapels in it (Prague)'

In (13-a), the subject of the main clause is non-thematic *it* and, in (13-b), it is non-thematic *there*.

Let us now go back to the German data discussed in Berman (2003). In German, it does not seem as if the passive raising construction, exemplified for English in (10), is grammatical. Consider the sentence in (14).

[9]Evidence that such an extra argument is not available in the active voice is for instance that we don't find sentences such as the one in (i).

(i) *He said her [to be the death of her husband].

(14) *Sie wird gesagt, [der Tod ihres Mannes gewesen zu sein].
She is said the death her man been to be
'She is said to have been the death of her husband' [constructed]

I take the ungrammaticality of (14) as an indication that the passive constructions in German and English are different. I will argue that *werden* is not a raising verb in German, and that only the *it*+ADJ construction is available for passive constructions in German. Consider the sentences in (7), here repeated as (15). Additional information has been added about the grammaticality of the sentences where there is no subject *es*.

(15) a. weil (es) gesagt wurde, [dass Hans krank ist].
because it said was that Hans sick is
'because it was said that Hans is sick.'
b. Was wurde (*es) gesagt, [dass er gelesen hat].
What was it said that he read has
'What was it said that he has read.'

Just like Berman (2003, 156), I take it that the ungrammaticality of extraction in (15-b), when a subject *es* is present, is a result of the fact that the *dass*-clause is an adjunct, out of which extraction is typically unacceptable. When no subject *es* is present, I assume that the subclause is demoted to complement, and that there is no subject in the clause. The argument-to-function mapping of passive *gesagt-werden* in German, applicable to the sentence without a subject *es*, is shown in (16).

(16) The argument structure of passive *gesagt-werden* with demotion of the propositional argument:

	agent	proposition
	|	|
gesagt-werden	⟨arg1	arg4⟩
	[–o]	[–o]
	[+r]	[+r]
	|	|
	(OBL$_{agent}$)	COMP

Just like in the argument-to-function mapping of English passive *said*, which was shown in (11), in (16), the first argument is demoted to OBL$_{agent}$ and the second argument is demoted to COMP.

For the sentence when a subject *es* is present, or for the preposed clausal subject, given in (17) from Berman (2003, 153), I assume that there is no demotion of the arg4[–o] argument, which results in the argument-to-function mapping in (18).

(17) [Dass Hans krank ist], wurde (*es) gesagt.
that Hans sick is was it said.
'That Hans is sick, was said.'

(18) The argument structure of passive *gesagt-werden* without demotion of the

propositional argument:

	agent	proposition
	|	|
gesagt-werden	⟨arg1	arg4⟩
	[–o]	[–o]
	[+r]	
	|	|
	(OBL_{agent})	SUBJ

In contrast to the mapping in (16), in (18), there is no demotion (adding a [+r] feature) of the propositional argument. The ungrammaticality of a subject *es* in the preposed clausal subject construction in (17) follows from the assumption that the preposed clausal subject is not demoted to COMP. Given the fact that there thus already is a subject in the clause, the addition of *es* is ungrammatical, violating the principle of coherence (Bresnan et. al., 2016).

Berman (2003, 165) gives a different analysis of (17), where the subclause does not constitute a subject and where there is a '(dropped) resumptive pronoun' blocking the presence of a subject *es*. The assumption that there is an empty pronoun blocking the presence of *es* seems unnecessarily complicated. Furthermore, it is not made clear why the resumptive pronoun sometimes is present covertly, while the subject pronoun *es* is never given the same analysis.

3.2 The *it*+ADJ *construction*

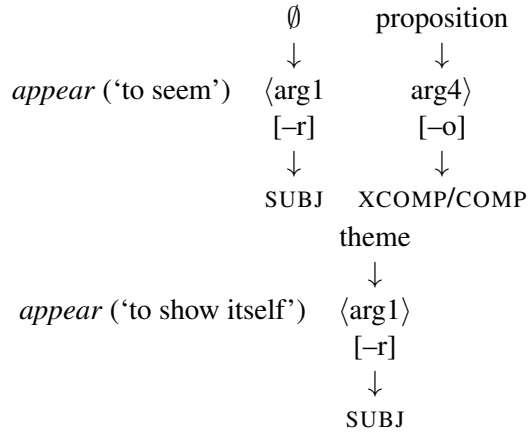
As mentioned above, the *it*+COMP in English is only used in connection with raising verbs, including raising *be*. With respect to non-raising predicates, a subject *it* is always analyzed as thematic. Consider for instance the analysis of the verb *appear*, which has two different lexical entries corresponding to two distinct interpretations, one of which is connected to its use as a raising verb and one of which is connected to its use as a non-raising verb. These two options are exemplified in (19).

(19) a. And a vision appeared to Paul in the night:
'And a vision showed itself to Paul in the night.'
(AUTHNEW-E2-P2,16,1A.1072)

b. The children appeared to be struck with amazement,
'It seemed that the children were struck with amazement.'
(COOK-1776,29.535)

In (19-a), the verb *appear* has the approximate interpretation 'to show itself' and functions syntactically as an intransitive verb. In (19-b), the same verb has the approximate interpretation 'to seem, to give the impression of being'. With this second interpretation the verb functions as a raising verb. The two argument structures for the two lexical entries for the verb *appear* are represented in (20).

(20) The two argument structures of *appear*:



The two different argument structures of the verb *appear*, given in (20), explain the differences in grammaticality between the two sentences in (21) and (22) below. The sentences in (21-a) and (22) are attested in the Late and Early Modern English corpora (Kroch et al., 2005, 2010), respectively, while the sentence in (21-b) is constructed. The ungrammaticality of sentences such as the one in (21-b) is frequently reported with respect to Present-Day English (e.g. Seppänen and Herriman, 2002), a judgement further supported by the fact that such sentences are not found in the corpora.

(21) a. - So it appears [to be].
(BOETHRI-1785,119.197)
b. *[To be so] appears. [constructed]

(22) a. [that in this matter I was not led by hym], very well and plainly appereth, (MROPER-E1-P1,521.98)
b. - it plainly appeared by this time [that he had got a stiff neck, as he never once more turned]. (COLLIER-1835,13.370)

In (21), the interpretation of *appear* is ‘to seem, to give the impression of’. With this interpretation, a subclause occurring in a preposed position on its own without a secondary predicate is ungrammatical, as illustrated in (21-b).

The fact that the propositional subclause in conjunction with raising verbs, including the verbs *appear* (‘to seem’) and *seem*, do not occur on their own has been referred to as obligatory extraposition (for further discussion see Seppänen and Herriman, 2002). One relatively recent account of obligatory extraposition, Alrenga (2005), connects the ungrammaticality of the structure in (21-b) to the complement selection of the verb. In particular, Alrenga (2005, 196) argues that a verb such as *seem* ‘only subcategorizes for a CP complement (seem: [_ CP])’. On my account, which differs from Alrenga (2005), the ungrammaticality instead follows from the principle of completeness (Bresnan et al., 2016), given the argument

structure assumed in (20). In (21-b), the verb *appear* ('to seem') does not have all the arguments it selects for. Given this analysis, no idiosyncratic selection for the syntactic category of the complement is required.

In (22), the interpretation of *appear* is 'to show itself'. With this interpretation, *appear* only selects for one argument, arg1[–r], which is mapped to SUBJ. This accounts for the fact that the sentence in (22-a) is grammatical, despite the fact that it superficially looks similar to the sentence in (21-b). In (22-b), where the subclause co-occurs with a subject *it*, we have the *it*+ADJ construction, which is the analysis given for all non-raising predicates, where a subject *it* occurs in conjunction with a propositional subclause.

A consequence of the analysis of a sentence as *it*+ADJ is that extraction out of the subclause should not be possible, or at least be significantly worse than for extraction out of a complement. We would thus predict that the sentence in (23), a variant of (22-a), is significantly worse than the sentence in (24-b). This seems to be the case.

(23) *By whom does it appear well and plainly [that I was not led in this matter]? [constructed]

(24) a. It appears [that Cobham took Raleigh to be either a God, or an Idol]. (RALEIGH-E2-P1,1,213.46)
b. What does it appear [that Cobham took Raleigh to be]? [constructed]

When *appear* in (23) is given the interpretation 'to show itself', the sentence in (23) seems to be considerably worse than the sentence in (24-b), where it has the interpretation 'to seem'.

With respect to the German example in (8-b) with the verb *stören*, the possibility of extraction with a subject *es* would not be expected. Since *stören* is not a raising predicate, we would expect the subject *es* to be analyzed as thematic and the subclause as an adjunct, out of which extraction is ungrammatical. However, consider the sentence in (25), taken from Alejchem (1922).

(25) Was stört es dich, [dass das Kind spielt]?
what bothers it you that the child plays
'What does it bother you that the child plays?'

The sentence in (25) has two interpretations. On the first interpretation, *was* functions as the complement of *spielen* in the subclause, while, on the second exclamative interpretation, it does not. The actual interpretation in the text is the second one, where *was* is not the complement of *spielen*. In a search for the string 'Wen stört es' in a subcorpus within the *Deutsches Referenzkorpus* (http://www1.ids-mannheim.de/kl/projekte/korpora/), all (27 out of 27) instances represent this second exclamative interpretation. If sentences such as (25) consistently are interpreted as non-extractions, this gives support to the hypothesis that extractions are dispreferred in this type of structure due to the fact that the subclause constitutes

an adjunct.

While the dubious grammaticality of the sentence with a subject *es* follows from the status of the subclause as an adjunct, the ungrammaticality of extraction without a subject *es* follows from the fact that the subclause then constitutes a subject.

In this section, I have argued for the adoption of two different analyses for a subject *it* in conjunction with a propositional subclause in English. In the analysis termed the *it*+ADJ construction, the subject *it* is thematic and the subclause an adjunct, while, in the analysis termed the *it*+COMP construction, the subject *it* is non-thematic and the subclause is a complement. In the next section, the question is asked whether the two constructions always have constituted part of the English language. As will be seen, there is evidence to suggest that the *it*+COMP construction was not part of the grammar of Old English.

4 *It*+ADJ and *it*+COMP in early English

As mentioned in the introduction, the alternation between the clausal subject construction and *it*+*subclause* construction seems to have been a part of the English language in all historical periods. For Old English, consider the sentences in (26). The examples come from the York-Toronto-Helsinki Parsed Corpus of Old English Prose (Taylor et al., 2003).

(26) a. Gregorius cwæð, on sumum timan gelamp, [þæt sum man forlet
Gregory said at some time happened that some man lost
his eagena gesihðe].
his eyes' sight
'Gregorius said that it happened at one time that some man lost his eyesight.'

b. Hit gelampt þa sume dæge [þæt Iosep wæs ana on his
It happened then some day that Joseph was alone in his
hlafordes huse].
master's house
'It happened one day that Joseph was alone in his master's house.'

The sentence in (26-a) contains the clausal subject *þæt sum man forlet his eagena gesihðe* as the only argument of the verb *gelimpan*. The sentence in (26-b), containing the same verb, *gelimpan*, has a subject *hit* ('it') in conjunction with the propositional subclause *þæt Iosep wæs ana on his hlafordes huse*. In contrast to Present-day English, clausal subjects in Old English most often occur in a clause-final position, rather than in a clause-initial position[10]. For support for the analysis of clausal-final subclauses of intransitive verbs as subjects, see Méndez Naya

[10]In the Old English prose corpus, there are four clausal subjects that do not occur in a clause-final position, one clause-initial *wh*-clause, two clause-initial infinitival phrase and one clause-medial *that*-clause.

(1997) and Zimmerman (2015).

In the Old English prose corpus, there are 705 instances of a subject *hit* in conjunction with a prepositional subclause. The question to be discussed in the present section is whether sentences such as the one in (26-b) are to be analyzed as *it*+ADJ or *it*+COMP. Is the subject *hit* thematic or non-thematic? Does the subclause constitute an adjunct or a complement?

The discussion in the following will be based on data from the two verbs that most frequently occurs with a propositional subclause and without an agentive subject in the Old English corpus, namely *þyncan* ('to seem, appear') and *gelimpan* ('to happen'). The verb *gelimpan* alone accounts for about a third of the instances of the *it+subclause* construction in the Old English corpus (228 out of 705 instances). With 112 instances, the verb *thyncan* is also very frequent in the corpus.

The structure of the following discussion is as follows. First, the proposed argument structures for the verbs *þyncan* and *gelimpan* are given. Then, three types of support are given for these argument structures, concerning extraction data, the (non-)existence of raising, and co-occurence patterns between a subject *hit* and a dative experiencer.

The proposed argument structures for the verbs *þyncan* and *gelimpan* are given in (27) and (28), respectively. Two argument structures are given for each verb. The choice of between argument structures is assumed to follow from the mapping between thematic roles and argument slots, something that will not be further discussed here. An important part of the analysis of these verbs concerns the notion of dative subjects. As is discussed in Allen (1995), there is evidence, for instance from conjunction reduction, supporting the analysis of certain dative phrases in conjunction with intransitive verbs as subjects. More support for this analysis is given below.

(27) Argument structures of *þyncan*:

a. With dative subject (*me*+*verb*+[*that* . . .]):

	experiencer	proposition
	|	|
þyncan	⟨arg1	arg4⟩
	[–o]	[–o]
	|	|
	SUBJ	COMP

b. With dative object (*it*+*verb*+*me*+[*that* . . .]):

	proposition	experiencer
	|	|
þyncan	⟨arg1	arg3⟩
	[–r]	[+o]
	|	|
	SUBJ	OBJ

In (27), the argument structures for the verb *thyncan* is given. In (27-a), *þyncan* takes two arguments, arg1[–o] and arg4[–o], which are mapped to SUBJ and COMP, respectively. In (27-b), *þyncan* also takes two arguments, but now they are arg1[–r] and arg3[+o], which are mapped to SUBJ and OBJ, respectively.

The argument structures for the verb *gelimpan* are given in (28).

(28) Argument structures of *gelimpan*:

a. With dative subject (*me+verb+[that . . .]*):

	experiencer	proposition
	|	|
gelimpan	⟨arg1	arg4⟩
	[–o]	[–o]
	|	|
	SUBJ	COMP

b. Without experiencer argument (*(it)+verb+[that . . .]*):

	proposition
	|
gelimpan	⟨arg1⟩
	[–r]
	|
	SUBJ

In (28-a), the argument structure is the same as the one shown for *þyncan* in (27-a). In the second one, the verb *gelimpan* takes just one argument, arg1[–r], which is mapped to SUBJ.

Let's proceed to the data that support the adoption of the argument structures in (27) and (28). First, let's consider the co-occurrence between a subject *hit* ('it') and a dative experiencer for *þyncan* and *gelimpan*, when they occur with a propositional subclause. The frequencies are shown in Table 1. The frequences are per 100,000 clauses[11], with the token frequency within parenthesis.

Table 1: The verbs *þyncan* and *gelimpan* in conjunction with a clausal argument and the pronoun *(h)it* and/or an oblique experiencer per 100,000 clauses.

	both *hit* and dative experiencer	*(h)it*	dative experiencer	neither *hit* nor dative experiencer
þyncan	<1 (2)	0 (0)	46 (109)	<1 (1)
gelimpan	0 (0)	97 (228)	13 (30)	27 (63)

In Table 1, it is shown that the verb *þyncan* more or less consistently occurs together with only an experiencer argument and the subclause. This supports the

[11] In the YCOE, all clauses are represented as IPs. The corpus contains 236,046 IPs

argument structure in (27-a), where *thyncan* takes two arguments, a dative experiencer subject and a propositional complement (the subclause). However, there are also two cases in which *þyncan* occurs with both a dative experiencer and a subject *hit*. One of these two sentences is given in (29).

(29) Wel geradlic hyt eac þingð us þæt we herto gecnytton þa epactas,
well appropriate it also seems us that we hereto tied those epacts
'It seems very appripriate to us that we tied the epacts to this'.
(cobyrhtf,ByrM_1_[Baker-Lapidge]:1.2.291.403)

In (29), *þyncan* takes both a subject *hit* and a dative experiencer, which is here analysed as an object. This is an *it*+ADJ construction. The argument-to-function mapping for the two sentences where *hit* and a dative experiencer co-occur is the one in (27-b).

For the verb *gelimpan*, the most common occurrence is when *gelimpan* occurs with a subject *hit*, without a dative experiencer, or when there is only a propositional subclause, with neither experiencer nor *hit*. This gives support for the argument structure in (28-b), where there is only one argument, the propositional argument. The most frequent alternative, where *gelimpan* only occurs with *hit* and the subclause represents the *it*+ADJ construction. However, there is also a considerable number of sentences where *gelimpan* occurs with a dative experiencer instead of a subject *it*. This gives support for the argument structure in (28-a), where *gelimpan* has the same argument structure as *thyncan* in (27-a). For the verb *gelimpan*, in the data shown here, there is complementary distribution between a subject *hit* and a dative experiencer. This can be taken as evidence for the argument structure where the dative experiencer constitutes SUBJ.

For the verb *thyncan*, there is no real complementary distribution between dative experiencers and a subject *hit*. The need for the dative subject analysis for *þyncan* is thus not clear from Table 1. However, one additional piece of support for subject analysis of the dative experiencer comes from extractions. When a dative experiencer co-occurs with a *that*-clause, it seems to be possible to extract out of the *that*-clause. One example is given in (30).

(30) Hwæt þincð þe [þæt þu sy]?
what seems you that you be
'What do you think you are?'
(cowsgosp,Jn_[WSCp]:8.53.6483)

In (30), the *wh*-phrase *hwæt* is extracted out of the *that*-clause, giving support to the argument structure in (27-a), where the dative experiencer is SUBJ and the propositional argument COMP. If the subclause constitutes a subject, we would not expect extraction to be possible here.

As can be seen from the argument structures in (27) and (28), *þyncan* and *gelimpan* are given different analyses than the raising verbs *seem* and *appear* in more modern English. As discussed in section 3, the *it*+COMP construction, where a non-thematic subject *it* occurs in conjunction with a complement subclause, is connected to the analysis of the verb as a raising verb. If the *it*+COMP construction were a part of Old English grammar, we would expect typical raising sentences, exemplified in (10), to alternate with the constructions represented in (26) in Old English.

With respect to the verb *þyncan*, whose argument structures is given in (27), there are 15 instances in the Old English prose corpus, where this verb occurs together with an infinitival phrase. In 14 out of 15 sentences, a thematic argument of the main clause is also the subject of the infinitival phrase, i.e. it is control rather than raising that gives the identification of the subject of the infinitival phrase. There is, however, one sentence that seems to represent raising. This sentence is gvien in (31).

(31) swa þæt me þynceþ [of gemynde beon] Paulines wundor Nolane burge
so that me seems of memory be Paulinus' miracle Nola city
biscopes,
bishop
'so that the miracle of Pauline, bishop of the city of Nola, seems to be forgotten'.
(cogregdC,GDPref_and_3_[C]:0.179.4.2177)

In (31), the nominative phrase *Paulines wundor* constitutes the subject of both the verb *þyncan* and the infinitival phrase *of gemynde beon*. The example in (31) should thus be analyzed as raising and constitutes a counterexample in relation to the argument structures given in (27). However, as discussed in Denison (1993), it is somewhat questionable whether the particular example given in (31) constitutes 'natural' Old English. As pointed out by Denison (1993, 221), the structure and word order of the sentence in (31), which is part of a translation from Latin, seems to follow the Latin original in a rigid manner[12].

The lack of raising structures for the verb *þyncan* constitutes an additional piece of support for the argument structures in (27), and for the fact that the *it+subclause* construction with the verbs *thyncan* and *gelimpan* should be analysed as *it*+ADJ, rather than *it*+COMP.

[12]The Latin original is given in (i).

(i) ita ut Paulini miraculum, Nolanae urbis episcopi, . . . , memoriae defuisse videatur.

5 Conclusion

This paper has given an analysis of the argument structure of predicates that alternately take a clausal subject, and a subject *it* in conjunction with a propositional subclause in present-day and historical English. Using the revised Lexical Mapping theory of Kibort (2007, 2008, 2013, 2014), two separate argument structures are given for constructions with a subject *it* in conjunction with a subclause. On the one hand, we have a thematic subject co-occurring with an adjunct subclause, the *it*+ADJ construction. On the other hand, we have a non-thematic subject *it* co-occurring with a complement subclause, the *it*+COMP construction. The presence of these two argument structures accounts for a number of differences between raising predicates and non-raising predicates when it comes to constructions they realize. For one thing, it straightforwardly provides an explanation why a subclause cannot occur as a preposed clausal subject without a secondary predicate with raising verbs. With respect to the historical English data, the absence of the *it*+COMP construction explains facts about extraction and co-occurrence patterns for verbs that occur together with a propositional subclause.

References

Alejchem, Scholem. 1922. *Aus dem nahen Osten*. Benjamin Harz Verlag.

Allen, Cynthia. 1995. *Case Marking and Reanalysis: grammatical relations from Old to Early Modern English*. Oxford: Oxford University Press.

Alrenga, Peter. 2005. A sentential subject asymmetry in English and its implications for complement selection. *Syntax* 8(3), 175–207.

Anderson, John M. 1997. Preliminaries to a history of sentential subjects in English. *Studia Anglica Posnaniensia XXXI* .

Berman, Judith. 2003. *Topics in the Clausal Syntax of German*. Stanford, CA: CSLI Publications.

Bresnan, Joan, Asudeh, Ash, Toivonen, Ida and Wechsler, Stephen. 2016. *Lexical-Functional Syntax*. Oxford/Malden (Mass.): Blackwell Publishers Ltd, second edition edition.

Denison, David. 1993. *English historical syntax: verbal constructions*. London/New York: Longman.

Haugland, Kari E. 2006. *Old English impersonal constructions and the use and non-use of nonreferential pronouns*. Ph. D.thesis, Universitetet i Bergen, Bergen.

Huddleston, Rodney D. and Pullum, Geoffrey K. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Kibort, Anna. 2007. Extending the applicability of Lexical Mapping Theory. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG07 Conference*, CSLI Publications.

Kibort, Anna. 2008. On the Syntax of Ditransitive Constructions . In Miriam Butt and Tracy Holloway King (eds.), *The Proceedings of the LFG '08 Conference*, pages 312–332, University of Sydney, Australia.

Kibort, Anna. 2013. Objects and Lexical Mapping Theory. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG13 Conference*, Stanford, CA: CSLI Publications.

Kibort, Anna. 2014. Mapping out a construction inventory with (Lexical) Mapping Theory. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG14 Conference*, Stanford, CA: CSLI Publications.

Kroch, Anthony, Santorini, Beatrice and Delfs, Lauren. 2005. Penn-Helsinki Parsed Corpus of Early Modern English, CD-ROM, first edition, (http://www.ling.upenn.edu/hist-corpora/).

Kroch, Anthony S., Santorini, Beatrice and Diertani, Ariel. 2010. Penn Parsed Corpus of Modern British English, CD-ROM, first edition, (http://www.ling.upenn.edu/hist-corpora/).

Méndez Naya, Belén. 1997. Subject clauses in Old English: Do they really exist? *Miscelánea. A Journal of English and American Studies* 18, 213–230.

Patejuk, Agnieszka and Przepiórkowski, Adam. 2014. In favour of the raising analysis of passivisation. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG14 Conference*, Stanford, CA: CSLI Publications.

Seppänen, Aimo and Herriman, Jennifer. 2002. Extraposed Subjects vs. Postverbal Complements: On the So-Called Obligatory Extraposition. *Studia Neophilologica* 74(1), 30–59.

Taylor, Ann, Warner, Anthony, Pintzuk, Susan and Beths, Frank. 2003. The York-Toronto-Helsinki Parsed Corpus of Old English Prose, oxford Text Archive, first edition, (http://www-users.york.ac.uk/ lang18/pcorpus.html).

Zaenen, Annie and Engdahl, Elisabet. 1994. Descriptive and theoretical syntax in the lexicon. In B. T. S. Atkins and Antonio Zampolli (eds.), *Computational Approaches to the Lexicon*, pages 181–212, Oxford: Oxford University Press.

Zimmerman, Richard. 2015. Early English clausal arguments of intransitive verbs: Subjects or associates of empty expletives? In *The York-Newcastle-Holland Symposium on the History of English Syntax 13 (SHES13)*.

THE DIACHRONY OF DATIVE SUBJECTS AND THE MIDDLE IN ICELANDIC: A CORPUS STUDY

Christin Schätzle
University of Konstanz

Miriam Butt
University of Konstanz

Kristina Kotcheva
Universität Freiburg

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King (Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

Icelandic is very well known for non-nominative subjects. In recent years, it has been proposed that dative subjects are a Proto-Indo European feature, and that a Dative Subject Construction has been inherited through the ages in the daughter language families (Barðdal and Eythórsson, 2009; Barðdal et al., 2012). We conduct a corpus study and show that while dative subjects can indeed already be found in the earliest attested Icelandic texts, their distribution has been changing over the last millenium. In particular, their use in middles has increased significantly. We explain our findings via an increased use of experiencer subjects combined with a more regular association of experiencer arguments with dative case. We provide a formal analysis within LFG's Mapping or Linking Theory that draws on Barron's (2001) analysis of the diachronic development of raising verbs in Latin. Overall, we see our work as providing evidence against dative subjects in Icelandic as being due to an inherited monolithic Dative Subject Construction.

1 Introduction

Dative subjects are common in a variety of modern Indo-European languages. In recent years, their origin has been discussed controversially. The so-called Oblique Subject or Semantic Alignment Hypothesis (see, e.g. Barðdal and Eythórsson 2009; Barðdal et al. 2012) takes dative subjects to be a common Proto Indo-European inheritance. In contrast, the so-called Object-to-Subject Hypothesis assumes that dative subjects were innovated at a later stage in that former dative objects were reanalyzed as subjects (see, e.g. Haspelmath 2001). The latter is the more standard hypothesis and is supported by the empirical facts in Indo-Aryan, where no evidence for dative subjects can be adduced for Old Indo-Aryan (Hock, 1990), but where there is evidence for objects being reanalyzed as dative subjects in several stages (Deo, 2003; Butt and Deo, 2013). The conditioning factors involve verbal lexical semantics and the increasing systematic association of dative case with experiencers arguments.

The oldest attested texts for Icelandic stem from the 12th century. This is about the time that new case markers entered the system in Indo-Aryan (the old system having eroded over the course of Middle Indo-Aryan). It is from the 11th to 12th centuries onwards that dative subjects begin to be possible in Indo-Aryan and evidence for continuiung change in progress can be found in Indo-Aryan.

Given the Indo-Aryan situation, we asked ourselves whether a similar time line might not be at work in Icelandic and whether evidence for a change in progress with respect to dative subjects could be found in Icelandic. We investigated this

[†]We would like to thank the audiences of LFG15 for valuable feedback and comments on this work. We would also like to thank the two reviewers and Jóhannes Gísli Jónsson, Anna Kibort, Joan Maling and Annie Zaenen, who took valuable time to engage with our paper and provide us with invaluable feedback. We would also like to acknowledge the DFG-funded project on *Visual Analysis of Language Change and Use Patterns* which supported Christin Schätzle's part of the research.

by conducting a corpus study of the distribution of dative subjects in the Icelandic Parsed Historical Corpus (IcePaHC; Wallenberg et al. 2011). Our aim was to identify factors conditioning the distribution of dative subjects in Icelandic.

Our results show that the use of dative subjects overall is on the rise and is strongly associated with the middle morpheme *-st*.[1] These dative subject middles are mainly found with psych predicates, raising predicates and some specialized lexicalized usages like 'seem', which originally meant 'like'. Nominative subjects are correspondingly decreasing.

The overall picture that emerges from the corpus study suggests that dative case is becoming more systematically associated with lexical semantic factors. This does not support the idea of dative subjects as a stable, common Proto-Indo European inheritance. Rather, the dative is part of a complex case system that encodes (lexical) semantic distinctions. Our data can be understood via Kibort's Mapping Theory (cf. Kibort 2013, 2014) in combination with Kaufmann's (2007) analysis of the middle and Barron's (2001) LFG analysis of the diachronic change of verbs of perception to raising verbs and epistemic modals. Our findings are also in line with Jónsson (2003), who argues that lexical semantics are the major conditioning factor for case in Icelandic and that Icelandic is seeing a change by which dative subjects are becoming systematically associated with experiencers and goals.

2 Icelandic: Basic Data

Icelandic is an accusative language, with nominative being the default subject case, and accusative the default for direct objects. Although nominative is the most common morphological case for subjects, the synchronic existence of non-nominative subjects is well established (cf. Andrews 1976, 2001; Zaenen et al. 1985). The word order of Icelandic is fairly fixed.

2.1 Non-nominative subjects in Icelandic

Non-nominative subjects, including dative subjects are attested in the earliest Icelandic texts, which date from the 12th century. Old Scandinavian, the ancestor language of Icelandic, is said to display a high degree of correlation between morphological case and semantic role (Faarlund, 1994, 59). Nominative case was characteristic of the agent role, dative was typically used for marking the semantic roles of instrumental, recipient and ablative. Accusative case was associated with themes, and genitive with partitives (in a wider sense). There was no one-to-one relation of semantic role and morphological case. Both subjects and (direct) objects could be in any of the four cases (Thráinsson 1994, 175; Barðdal 2001).

In Modern Icelandic, all non-nominative subjects are said to be non-agentive (Thráinsson 1994, 176, Jónsson 2003). Dative subjects have been associated broadly

[1] Wood (2015) analyzes this as a clitic, albeit as one with reduced mobility. The precise morphological status of *-st* is not relevant for the discussion in this paper.

with two major classes of verbs: 1) experiencer or psych verbs; 2) verbs dubbed "happenstance" by Barðdal (2011), which include verbs of gain/success, happening, hindrance, ontological states, speaking, possession, evidentiality and modals. This second class can also be seen as involving a type of experiencer/goal.

The overall pattern is illustrated in (1) for the experiencer verb *líka* 'to like' (1) from the oldest text in the IcePaHC called "Fyrsta málfræðiritgerðin" (The First Grammatical Treatise), dated around 1150 CE.

(1) Vel líkuðu goðrøði góð røði, [...]
well like.PAST.3.PL G.DAT.SG good.NOM.PL oar.NOM.PL
'Goðrøði (the good oarsman) liked good oars well, [...]'

Verb agreement is with the nominative 'oars' and not with the dative 'Goðrøði'. Non-nominative subjects pattern like nominative subjects with respect to subjecthood characteristics. A notable exception is subject-verb agreement. The verb only agrees with nominative subjects (Zaenen et al. 1985; Thráinsson 1994, 176).

2.2 Dative Substitution or "Dative Sickness"

Icelandic is currently undergoing a change in case marking which began in the latter part of the 19th century. This "Dative Substitution" or "Dative Sickness" is a process by which accusatives are replaced systematically by datives, as shown in (2) where the original accusative subject in (2-a) is replaced by a dative in (2-b). Smith (1996) and Jónsson (2003) describe this as a change in progress in which datives systematically become associated with experiencers in Modern Icelandic.

(2) a. Mig langar að fara.
I.ACC long.PRES to go
'I long to go.'

b. Mér langar að fara.
I.DAT long.PRES to go
'I long to go.' (Smith, 1996, 22)

Dative Sickness points to lexical semantics as a conditioning factor for dative subjects. However, it does not say much about whether dative subjects are an inherited construction or whether there is evidence for structural change.

2.3 Dative Subjects and Voice

Icelandic distinguishes three voices: active, passive and middle. A nominative subject and an accusative object constitute the regular arguments of transitive active clauses. Under passivization the accusative object is realised as a nominative subject, as seen in (3). Passive is formed periphrastically via the copula *vera* 'to be' in

conjunction with a past participle.[2]

(3) a. einhver barði strákana í skólanum
somebody.NOM hit.PAST.3.SG the.boy.ACC.PL in the.school.DAT.SG
'Somebody hit the boys in school.'

b. strákarnir voru barðir í
the.boy.NOM.PL be.PAST.3.PL hit.PPART.M.NOM.PL in
skólanum
the.school.DAT.SG
'The boys were hit in school.' (Thráinsson, 1994, 177)

In contrast, dative and genitive objects generally preserve their case marking under passivization, as shown in (4) for a dative.[3]

(4) a. Skipstjórinn sökkti skipinu.
the.captain.NOM.SG sink.PAST.3.SG the.ship.DAT.SG
'The captain sank the ship.'

b. Skipinu var sökkt af
the.ship.DAT.SG be.PAST.3.SG sink.PPART.N.NOM.SG by
skipstjóranum.
the.captain.DAT.SG
'The ship was sunk by the captain.' (Zaenen and Maling, 1984)

Besides active and passive, Icelandic verbs can be marked for middle voice via the suffix *-st* (Sigurðsson, 1989; Anderson, 1990; Wood, 2015), as shown in (5).[4] As with passives, the accusative object of a transitive verb in an active clause is realised as a nominative subject in its middle counterpart.

(5) a. Ég fann hestinn.
I.NOM find.PAST.1.SG the.horse.ACC.SG
'I found the horse.'
b. Hesturinn fannst.
the.horse.NOM.SG find.PAST.MID.3.SG
'The horse got found.' (Sigurðsson, 1989, 243)

Dative objects vary with respect to middle formation. When the object is a dative theme/patient as in (4-a), the dative case is not preserved, as shown in (6). When the dative marks benefactives or goals, as in (7), the dative is preserved.

[2]The past participle agrees in number and gender with the nominative subject of the passive clause, but not with non-nominative subjects as per the general rules of Icelandic verb agreement, where they display invariant third person singular agreement (Thráinsson, 1994, 177).

[3]Prepositions select for case and the 'by' in passive by-phrases selects for a dative.

[4]This distribution of the Icelandic *-st* morphology is complex and not all of the verbs carrying *-st* are middles. However, the great majority are compatible with middle meanings as identified crosslinguistically (Kemmer, 1993; Kaufmann, 2007). Also see Wood (2015).

(6) a. Þeir sökktu bátnum.
they.NOM sink.PAST.3.PL the.ship.DAT.SG
'They sank the ship.'
b. Báturinn sökk.
the.ship.NOM.SG sink.PAST.3.SG
'The ship sank.' (Sigurðsson, 1989, 277)

(7) a. Pétur bauð mér vinnu.
Peter.NOM offer.PAST.3.SG I.DAT job.ACC.SG
'Peter offered me a job.'
b. Mér bauðst vinna.
I.DAT offer.PAST.MID job.NOM.SG
'I got the opportunity to get a job.' (Sigurðsson, 1989, 260)

Wood (2015) argues that datives as in (7) share characteristics with dative goal objects in ditranstives and should be analyzed as applied objects added to a predication via a specialized functional head (Appl). While building on Wood's basic insights, we provide an alternative analysis of (6) and (7) in section 4.

3 Corpus Study

The goal of our corpus study was to investigate whether dative subjects as a construction have been stable over time in Icelandic or whether evidence for structural change could be found. We based our study on the IcePaHC, which includes texts from the earliest attested stages of Icelandic up to modern times, i.e. from the 12th to the 21st century. There are 60 texts from different genres (mainly Sagas) containing approximately one million words. The corpus is annotated syntactically in the Penn Treebank style (Marcus et al., 1993), which includes the annotation of case and grammatical relations. As a sample, we provide the annotation of (8) in Figure 1. The sentence contains a dative subject with the predicate *finna* 'seem'.

(8) Mér finnst sem ég sé
I.DAT seem.PRES.MID.SG like I.NOM be.SBJNC.PRES.1.SG
sloppinn úr einhvers konar
escape.PPART.M.NOM.SG from some.GEN.SG kind.GEN.SG
fangelsi.
prison.DAT.SG
'It seems to me as if I have escaped from some kind of prison.'
Brynjólfur Sveinsson biskup, 1882

As can be seen, there is a matrix IP with a pronominal dative subject (*mér* 'I.DAT'). The verb is an inflected middle form: *finnst*. The annotation also provides a lemma, in this case the infinitive form *finna*. The verb embeds a CP that contains an IP with a pronominal nominative subject, and so on, building a hierarchical constituency and dependency structure.

```
(IP-MAT-SPE (NP-SBJ (PRO-D Mér-mér))
    (VBPI finnst-finna)
    (CP-ADV-SPE (WADVP-1 0)
      (C sem-sem)
      (IP-SUB-SPE (ADVP *T*-1)
                (NP-SBJ (PRO-N ég-ég))
                (BEPS sé-vera) (VBN sloppinn-sleppa)
                (PP (P úr-úr) (NP (NP-POS (ONE+Q-G einhvers-einhver)
                (N-G konar-konar)) (N-D fangelsi-fangelsi)))))
      (. .-.))
  (ID 1882.TORFHILDUR.NAR-FIC,.603))
```

Figure 1: Sample annotation from IcePaHC

3.1 Corpus Study Part I — Conditioning Factors

The goal of the first part of our corpus study was to identify factors which condition the appearance of dative subjects in Icelandic. We examined the interaction between dative subject case and factors which have previously been implicated in the literature (e.g. Barðdal et al. 2012; Barðdal and Eythórsson 2009; Barðdal 2011; Svenonius 2002; Maling 2002; Svenonius 2006; Jónsson 2003; Jónsson 2013), namely semantic verb class and voice in all sentences of the corpus.[5]

In order to investigate whether verb class is a significant conditioning factor, we manually annotated the IcePaHC with additional information about semantic verb classes. We based our annotations on Levin's classification for English (Levin 1993) in combination with previous literature on verb classes in Icelandic (Barðdal et al. 2012). An example of this augmented annotation is provided in Figure 2, in which the verb *finna* is classified as a psych verb (PSY).[6]

```
(IP-MAT-SPE (NP-SBJ (PRO-D Mér-mér))
    (VBPI-PSY finnst-finna)
            ....
```

Figure 2: Sample annotation of (8) with additional verb class information.

In accordance with the literature on Icelandic (e.g. Haugen 1984), we divided our data and results according to the time stages displayed in Table 1, which shows

[5]In our study, we took the syntactic annotation in the corpus at face value. However, we did find instances of erroneous annotation. We were able to identify some of these as part of the corpus study and duly disregarded them. Others were identified in retrospect by two critical readers: Joan Maling and Jóhannes Gísli Jónsson. Erroneous annotations include, for example, dative subjects on ditransitives and accusative subjects in conjunction with middle morphology, both of which did not and do not exist in Icelandic. We have been in touch with the corpus developers and the errors brought out by our study have now been corrected.

[6]The following categories for semantic verb classes were used in the corpus study: psych, sending and carrying, exerting force, appearance, communication, change of state, combining and attaching, involving the body, judgement, change of possession, removing, verbs with predicative complements, existence, perception , motion, aspect, desire, putting, measure, destruction, social interaction, throwing, ingestion, lingering and rushing, creation and transformation, emmission, killing, concealment, search, cutting, posture, contact by impact, learning, poking, image creation, lodging, and cognition. These verb classes are too fine grained and will be revised as part of future work.

our results with respect to voice. We used χ^2 to test whether the observed distributions differ from what could be expected (* p<0.05, ** p<0.01, *** p<0.001).

Time	active	passive	middle	χ^2
1150-1350	64.4%	18.2%	17.4%	***
1350-1550	66.8%	17.5%	15.7%	***
1550-1750	46.1%	28.8%	25.1%	***
1750-1900	53.1%	20.8%	26.2%	
1900-present	43.2%	14.3%	42.5%	***
all	55.3%	19.5%	25.5%	

Table 1: Diachronic distribution of dative subject predicates by voice

The data confirm that voice is relevant. There are changes in the frequency of actives, passives and middles occurring with dative subjects over time. Dative subjects appear most often with actives, but also fairly often with middle morphology. Dative subjects in active constructions are mostly experiencers and are mainly found with psych predicates (mean=53.4%). An illustrative example is given in (9).

(9) Það líkaði biskupi illa.
this.NOM.SG please.PAST.3.SG bishop.DAT.SG badly
'The bishop disapproved of this.' *Brynjólfur Sveinsson biskup, 1882*

Over the whole time span, dative subjects are found less often with passives. These passive dative subjects are mainly associated with verbs of communication (mean=14.8%), e.g. *tilkynna* 'announce' in (10), and change of possession (mean=14.6%) as *kaupa* 'buy' in (11), but are also decreasing within these verb classes over time in keeping with the overall decrease of passive dative subjects.

(10) en drottningu var tilkynnt að
and queen.DAT.SG be.PAST.3.SG announce.PPART.N.NOM.SG that
Þorsteinn væri afturkominn.
Þ.NOM.SG be.PAST.SUBJ.3.SG return.PPART.M.NOM.SG
'And it was announced to the queen that Þorsteinn had returned.'
Söguþáttur af Ármanni og Þorsteini gála, 1675

(11) Var þessu nú keypt.
be.PAST.3.SG this.N.DAT.SG now buy.PPART.N.NOM.SG
'This was now bought.' *Gunnars saga Keldugnúpsfífls, 1400*

While dative subjects with actives and passives show an overall decrease, middle morphology with dative subjects is on the increase. Dative subjects with middles also occur most often with psych predicates (56.2%), as in (12).

(12) Jú, mér leiðist ósköp, segir Sigríður.
yes I.DAT.SG be.bored.PAST.MID.SG much say.PRES.3.SG S.NOM.SG
'Yes, I am very bored, Sigríður said.' *Piltur og stúlka, 1850*

Note that the data in the table would seem to indicate that significant changes happened in the third time stage (1550–1750). However, as discussed in other work (Butt et al., 2014), these deviating percentages are due to a genre effect. While Sagas predominate in the other time stages, the third time stage is mainly represented via religious and legal texts in IcePaHC. The genre effect is clearly visible in the glyph visualization shown in Figure 3. We also used this visualization to explore the interaction between semantic verb class and dative subject marking.

3.2 Semantic Verb Classes

Since lexical semantics plays a role in the appearance of dative subjects, we are investigating the interaction between semantic verb class and dative subjects. Our data has so far yielded some suggestive evidence, but no significant results.

Figure 3 shows a visualization of the entire IcePaHC. The texts are ordered chronologically on the vertical axis. The texts are arranged horizontally according to genre. Narrative sagas are to the left, the religious and legal texts more towards the center and the right. As can be seen, there is a time stage in which the religious and legal texts predominate. This is where the genre effect skews the data.

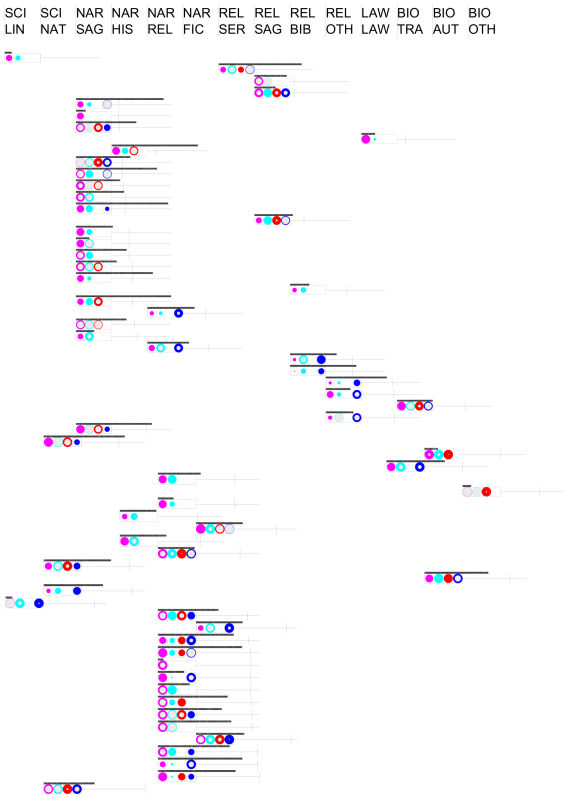


Figure 3: Multifactorial visualization of dative subjects in IcePaHC

Figure 4 displays one glyph from Figure 3. Each glyph represents one of the texts within IcePaHC. On top of the glyph is a horizontal bar which visualizes the text length. The horizontal bar consists of light and dark gray stripes. Each stripe represents a sentence. Light gray stripes indicate those sentences which contain a dative subject. Mousing over the gray stripes allows one to see the sentence involved. The circles represent the appearance of dative subjects with different verb classes. The semantic verb classes are as per the umbrella classifications in Barðdal et al. (2012) and the higher categorization which she assumes in Barðdal (2011).[7] They can be expanded into subclasses per mouse click, as shown in Figure 4. The circular glyphs are redundantly coded by color and indicate whether a given verb class or category appears more or less often than expected based on the text length and with respect to the entire corpus. If the circular glyphs are filled in from the outside, the given class or category occurs more often than expected. If the glyph is filled from the inside out, the given class or category occurs less often than expected (Schätzle et al., 2014).

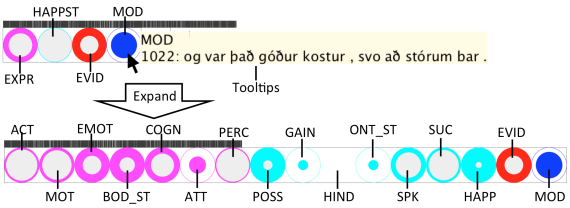


Figure 4: Text glyph visualization of dative subjects in a text from IcePaHC according to lexical semantic verb class and category

The visualization follows an 'overview first — details on demand' approach (Keim et al., 2008) allowing for interactive exploratory access to the data which is particularly useful for the analysis of a multifactorial diachronic data set. Interactive visualizations such as the one in Figure 3 are also good in situations where a purely quantitative approach to data is not fine-grained enough because there are too few instances per category to allow for firm statistical conclusions.

The visualization shows a preponderance of dative subjects in experiencer verbs (pink) as well as "happenstance" verbs (light blue), which include verbs of possession and gain and communication (SPK). The other major categories in evidence are verbs of evidentiality (dubbed thus by Barðdal; red) and modals (dark blue), which appear more frequently in the latter stages of the language.

[7]They also include verbs of motion and activity verbs. In future work, we will experiment with a different classification of verbs, in particular, we will experiment with broader umbrella categories, making it more likely that significant patterns can be found. Also note that Jónsson (2003) shows that further factors play a role. For example, verbs of emotion differ in that verbs of strong positive emotion never take a dative subject, but verbs of negative emotion do.

3.3 Corpus Study Part II — Voice

The first part of our study shows that dative subjects are on the increase in a subset of structural and lexical semantic contexts. Our results suggest that voice plays a major role in the diachronic development of dative subjects. In a second study, we therefore studied potential factors affecting the appearance of dative subjects with different voices. We quantitatively analyzed case, word order, and transitivity according to voice in all matrix declarative sentences in IcePaHC. Moreover, the following verb types in IcePaHC were analyzed separately: main verbs, modals, 'have', 'do', 'become', and 'be'. Again, we divided the data into the time stages suggested for Icelandic and conducted a χ^2-test for significance as before.

We first established the overall occurrence of dative subjects in comparison to other types of subjects. We extracted all subject cases in declarative matrix sentences from the corpus. We identified 65568 declarative matrix sentences in which the subjects were distinctly case marked by the annotation format of the corpus excluding ambiguous or non-marked subjects. Their diachronic distribution with respect to subject case is displayed in Table 2.

Time	NOM	ACC	DAT	GEN	χ^2
1150-1350	94.7%	0.9%	4.2%	0.2%	
1350-1550	96.1%	0.7%	3.1%	0.1%	***
1550-1750	95.0%	0.9%	4.1%	0.1%	
1750-1900	95.3%	0.6%	4.0%	0.1%	*
1900-present	93.1%	1.0%	6.0%	0.1%	***
all	95.1%	0.8%	4.2%	0.1%	

Table 2: Diachronic distribution of subject case

As expected, subjects are most often nominative. A small percentage of subjects bear dative case. Accusative and genitive subjects are rare. Diachronically, nominatives only decrease marginally, but significantly in the last time stage, while accusative and dative subjects increase highly significantly. The accusative subject numbers are small, so it is not clear whether this result should be weighed strongly.

This contrasts with Barðdal's (2008) corpus-based analysis of subject case in Old and Modern Icelandic based on texts coming from similar genres as the texts in IcePaHC. She shows that nominative subjects increase from 76.3% in Old to 85.0% in Modern Icelandic and that dative subjects decrease from 18.4% to 10.3% which clearly contradicts our findings. Accusative and genitive subjects are slightly reduced over time, but rarely appear overall. Additionally, the ratio of non-nominative subjects as opposed to nominative subjects is considerably higher in her study, i.e. 23.7%/15.0% to 76.3%/85.0% versus 4.9% to 95.1% in our study. However, Barðdal only considered subjects of transitive predicates while we did not distinguish between different valencies.

We then further identified 17523 matrix declarative sentences with distinctly case marked subjects and objects of transitive predicates. Our motivation was to

find out whether the case marking system found in the corpus corresponds to what is generally assumed about subject and object marking in Icelandic and to find out how dative case marked NPs in general develop over time. The distribution of object case in the history of Icelandic is shown in Table 3.

Time	NOM	ACC	DAT	GEN	χ^2
1150-1350	4.2%	72.0%	19.5%	4.3%	***
1350-1550	3.6%	69.7%	24.1%	2.6%	*
1550-1750	4.8%	71.3%	21.4%	2.5%	*
1750-1900	3.8%	67.3%	25.7%	3.2%	*
1900-present	4.9%	67.1%	25.6%	2.4%	***
all	4.2%	69.8%	22.9%	3.1%	

Table 3: Diachronic distribution of object case

As expected, objects are most often accusative, but dative objects are also common. Nominative and genitive objects appear much more infrequently. Genitive and accusative objects are slightly reduced over time, while dative objects increase significantly. Nominative objects stay fairly stable over the whole time span.

A data set examined by Barðdal (2001) yielded 93–94% nominative subjects, 67–69% accusative objects and about 25% dative objects, the numbers being stable in both historical and modern stages of Icelandic. These findings tally with our study. However, we do find evidence for change over time.

We examined subject and object cases individually according to different verb types. We found that verb type has no effect on the investigated factors case, voice, word order, and transitivity. Hence, we present our findings for main verbs in the following as they represent the majority of all verbs. We found 51209 matrix declarative sentences with case marked subjects and main verbs in the corpus. The distribution of case marking on subjects and objects for main verbs only is similar to the findings for all verbs (Tables 2 and 3). This is true for the distribution of subject case in active sentences as well (Table 2). However, passive sentences deviate from this overall pattern, see Table 4. Over time, dative subjects increase significantly in passives and nominative subjects decrease slightly.[8,9]

With respect to middles, as shown in Table 5, the increase in dative subjects is more marked and more significant.[10] Strikingly, nominative subjects are reduced

[8]Accusative subjects with passives are actually ungrammatical, see e.g. Svenonius (2006). The examples we found appear to be due to annotation mistakes, caused by the ambiguity of past and passive participles. Genitive subjects are indeed possible in passives, but are rarely used.

[9]The first part of our corpus study found that passives were decreasing, in particular with respect to two verb classes. The difference arises from the material under consideration. In this second part of the study, we looked only at matrix declaratives. In the first part, we also included participial constructions and small clauses.

[10]Further annotation mistakes were uncovered here. The genitive subjects in Table 5 in the third time stage should have been annotated as attributes, not subjects. Genitive subjects do not occur with middles, and neither do accusative subjects, these also represent annotation errors.

Time	NOM	ACC	DAT	GEN	χ^2
1150-1350	86.4%	0.3%	10.9%	2.4%	*
1350-1550	83.7%	0.0%	14.5%	1.8%	
1550-1750	83.3%	0.2%	16.0%	0.5%	
1750-1900	80.9%	0.0%	16.8%	2.3%	
1900-present	82.2%	0.8%	16.4%	0.6%	*
all	83.8%	0.3%	14.4%	1.6%	

Table 4: Diachronic distribution of subject case in passive sentences

by 17%, while middle morphology in the corpus is strongly associated with dative subjects, showing an increase of 17%. The middle verbs found in the corpus are mainly experiencer predicates, see e.g. (13), and raising verbs. As shown in (14), these are often based diachronically on former psych predicates or verbs of perception. In (14) the raising predicate is *finna* 'seem', which used to mean 'like' and still has a modern psych verb interpretation as 'feel/sense' in certain contexts.

Time	NOM	ACC	DAT	GEN	χ^2
1150-1350	92.8%	0.2%	7.0%	0.0%	***
1350-1550	94.4%	0.2%	5.4%	0.0%	***
1550-1750	92.3%	0.1%	7.5%	0.1%	*
1750-1900	89.5%	0.0%	10.5%	0.0%	
1900-present	75.5%	0.1%	24.4%	0.0%	***
all	89.3%	0.1%	10.6%	0.0%	

Table 5: Diachronic distribution of subject case with middles

(13) Mest af öllu gramdist honum það thó að sjá
most of all.DAT resent.PAST.3.SG he.DAT that still to see.INF
menn, sem ... bregðast sér
man.NOM.PL REL.PRO fail.PRES.MID self.3.SG.DAT
'But most of all it annoyed him to see men who ... fail him.'
Leysing, 1907

(14) Ekki finnast mér Sóróasters trúarbrögð eins
not seem.PRES.MID.SG I.DAT S.GEN.SG religion.NOM.SG one
herleg og þau sem Óðinn
glorious.M.NOM.SG and that.N.NOM.SG like Ó.NOM.SG
hafði og kenndi.
have.PAST.3.SG and teach.PAST.3.SG
'Zoroaster's teachings don't seem to me as glorious as the ones that Odin had and taught.' *Fimmbræðra saga, 1790*

As the increase in dative subjects is strongly correlated with a rise in middle

forms, we took a closer look at the interaction of dative subjects and middles in comparison with the interaction of dative subjects with other voices of main verbs in matrix declaratives. As shown in Table 6 and as expected, the use of dative subjects with middles increases over time. Additionally, active constructions appear less often with dative subjects.

Time	active	passive	middle	χ^2
1150-1350	70.4%	16.6%	13.0%	***
1350-1550	69.3%	16.4%	14.3%	***
1550-1750	49.2%	21.8%	29.0%	***
1750-1900	58.5%	24.3%	17.2%	
1900-present	41.0%	47.1%	11.8%	***
all	58.0%	25.9%	16.1%	

Table 6: Diachronic distribution of dative subject predicates by voice

Most of the dative subjects are found with intransitive verbs, see Table 7. Transitive predicates with dative subjects have the following mean object distribution: nominative 89.6%, accusative 6.7%, dative 2.2%, and genitive 1.5%.[11] With respect to word order, no significant deviances with respect to dative subjects from what was found for the whole corpus by Butt et al. (2014) were identified.

Time	intransitive	transitive	ditransitive	χ^2
1150-1350	72.3%	27.4%	0.3%	
1350-1550	69.0%	31.0%	0.0%	
1550-1750	68.5%	31.5%	0.0%	
1750-1900	78.3%	21.7%	0.0%	
1900-present	79.2%	20.8%	0.0%	*
all	73.8%	26.1%	0.1%	

Table 7: Diachronic distribution of dative subject according to transitivity

Our corpus study shows that dative subjects are on the increase overall. Additionally, there is a strong association of dative subjects with middle morphology. These middle forms are mainly found on psych and raising predicates. Nominative subjects with middles are correspondingly decreasing. Other possible factors, e.g. word order, modality, or verbal semantic class, are not significant.

[11] In fact, the number of genitives should be 0 as this combination is not possible, pointing to further annotation errors. The dative subjects found with ditransitives listed in Table 7 have also all been identified as annotation errors.

4 Analysis and Discussion

The corpus studies have yielded a rich set of observations that need to be understood further. However, a central result of our study is that dative subjects are on the increase overall in Icelandic and that the main factor governing this increase is middle morphology on verbs. In this section we concentrate on understanding this result and put forward a proposal that involves three major factors. One is an increasingly systematic association of dative case with experiencers in Icelandic (Smith, 1994; Jónsson, 2003). The second is the interaction between middle formation and the systematic association of experiencers and goals with the dative. The third is the development of new lexicalizations that instantiate dative experiencer verbs via middle formation and diachronic reanalysis.

4.1 Case and Thematic Role

In analyzing Dative Sickness in Icelandic, Smith (1996) describes it as a change in progress in which datives are becoming systematically associated with either goals or experiencers in Modern Icelandic. This Dative Sickness accounts for some of the increase in dative subjects in Icelandic. Our studies did not find this to be a significant factor with respect to its application to individual verbs. However, we hypothesize that in combination with middle formation, the systematic association of experiencers/goals with datives becomes potent.

4.2 Dative Objects

Objects in Icelandic exhibit a variety of case marking (Maling, 2002). As with other parts of the case marking system, the association between case marking and thematic roles or other event semantic factors is mostly regular (cf. discussions in Zaenen et al. (1985); Faarlund (1994); Jónsson (2003)). Additionally, Svenonius (2002) shows that dative and accusative objects alternate systematically in certain verbs whereby the dative is used when the external argument (the subject) is only part of the initiating or causing event, but does "accompany" the object to the result state. For example, verbs of ballistic motion in which a ball is thrown or an arrow loosed would take a dative object. On the other hand, when a ball is rolled alongside so that the roller and the ball end up in the same place, an accusative would be used. Svenonius' generalization can motivate the dative on objects as in (6) (the sinkers do not accompany the boat under the water).

Wood (2015) argues that dative objects can be either direct or applied objects. Examples such as (6) or (14) are instances of direct objects, datives as in (7) are applied objects. We adopt Svenonius' and Wood's basic insights and show how these can be used to understand the increase of datives with middle morphology.

4.3 Middle Formation and Datives

In providing an analysis of the middle, we draw on Kaufmann's (2007) analysis in conjunction with Kibort's (2013; 2014) revised Linking Theory. Kaufmann shows that in middles the agent argument is in principle present on a pragmatic or inferential level; however, it is not accessible for structural purposes. This analysis tallies well with the known facts for Icelandic by which middles are incompatible with agentive modifiers or by-phrases (Wood 2015, 139 and references therein). This stands in stark contrast to the passive, where the agent argument is still structurally available, it is just not realized as a subject, but as an OBL or an adjunct.

We posit the linking schema in Table 8 for middles. It shows that an agent is involved in the event in principle. However, this agent is not available for linking in the middle. The effect is that the agent is not structurally present.[12,13]

	agent		theme	
			|	
verb_{middle}	<		arg_2	>
			[-r]	
			|	
			SUBJ	
			NOM	

Table 8: Linking Schema for a Middle

When this linking schema applies to (6), the only argument for linking is the boat and it is realized as a nominative subject. A different situation obtains when the dative marks benefactive or goal arguments, as in (15).

(15) a. Pétur bauð mér vinnu.
Peter.NOM offer.PAST.3.SG I.DAT job.ACC.SG
'Peter offered me a job.'

b. Mér bauðst vinna.
I.DAT offer.PAST.MID job.NOM.SG
'I got the opportunity to get a job.' (Sigurðsson, 1989, 260)

If we adopt Wood's analysis, in examples as in (15), the dative is an applied argument. As per standard LFG (Dalrymple, 2001), applied arguments are treated as unrestricted [-r] arguments. This also fits in with the general observation that the goal argument in Icelandic is linked to the direct object in actives (Zaenen et al., 1985; Bresnan, 2001) and the accusative object is the secondary object (OBJ_θ).

As shown in Table (15), when middle formation applies, the agent is not available for linking. Instead, the applied [-r] object is linked to the subject. The remaining argument is linked to the direct object. Case marking here is sensitive to

[12] Wood (2015) essentially analyzes the middle *-st* as instantiating the agent argument.

[13] For a different LFG analysis of middles, see Arka (2015).

	agent	goal$_{appl}$	theme
		|	|
offer$_{middle}$	<	arg$_2$	arg$_3$ >
		[-r]	[+o]
		|	|
		SUBJ	OBJ
		DAT	NOM

Table 9: Linking Schema for the middle of 'offer'

various factors. For one, the subject is a goal and as such attracts dative case. For another, there is a constraint in Icelandic such that if the subject is non-nominative, then the object must be nominative (Jónsson, 2003).[14] The active secondary accusative object is thus realized as a nominative direct object in the middle.

Examples as in (15) add to the numbers of dative subjects found with middles. However, we suggest that the major factor governing the increase of dative subjects with middles is the formation of new experiencer predicates.

4.4 Verbs of Perception and Experiencer Predicates

A majority of the verbs with middle morphology and dative subjects in the corpus are experiencer predicates as in (16) (cf. also (8)). These experiencer predicates are generally cognate with a verb of perception, as shown in (17). A synchronic and diachronic relationship between experiencer predicates, including verbs like 'seem', and verbs of perception or feeling as in (17) can be observed crosslinguistically.

(16) Eiríki finnst þetta ómögulegt
Eric.DAT.SG find.PRES.MID.SG this.NOM.SG impossible
'Eric finds this impossible.' (Jónsson, 2003, 131)

(17) Haraldur fann það.
Harald.NOM.SG feel.PAST.3.SG it.ACC
'Harald felt/sensed it.' *Morkinskinna, 1275*

Barron (2001) proposes a diachronic LFG-based analysis for understanding the diachronic development from verbs of perception to experiencer and raising predicates. She identifies three constraints that must obtain for the diachronic development: 1) presence of secondary predication; 2) suppression of the perceiver argument; 3) shift from a physical to a mental process.

The shift from a physical to a mental process is given by pairs like (18) vs. (17). We propose that the second constraint is fulfilled by middle formation. (19) shows that *finna* can be used in contexts with secondary predication.

[14]There are some exceptions to this for restricted set of verbs involving either archaic usages or verbs with Acc-Acc patterns where the accusative subject is replaced with a dative via Dative Sickness (Wood, 2015, 45–46).

(18) Hann fann hestbein.
he.NOM.SG find.PAST.3.SG horse.bones.ACC.PL
'He found horse bones.' *Grettis saga Ásmundarsonar, 1310*

(19) og fann hertoginn að hann mundi skýr
and find.PAST.3.SG the.duke.NOM that he.NOM would intelligent.NOM
og málsnjallur vera.
and speech.excellent.NOM be.
'And the duke found that he would be intelligent and eloquent.'
Fimmbræðra saga, 1790

The examples in (17)–(19) all involve nominative subjects, but (16) has a dative subject. Working on Latin, Barron suggests that the dative perceiver/experiencer is first introduced as an optional argument in contexts where the nominative perceiver had been suppresed at argument structure (via passivization in her data). Although not realized syntactically, the perceiver is nevertheless part of the event semantics and can hence be introduced back into the syntax optionally. In Latin, this is achieved via a dative NP.

We suggest that essentially the same process is at work in Icelandic and that experiencers are (re)introduced in a clause via an applied dative. Consider (17). In middle formation, the agent and perceiver *Harald* is not expressed as per the linking schema in Table 4.3. However, the experiencer verb 'feel/sense' implies an experiencer/perceiver. This experiencer argument can be (re)introduced overtly via a dative applied object, as shown in Table (19).[15] The linking then proceeds exactly as in Table (15).

		agent	experiencer$_{appl}$	theme	
			|	|	
finna$_{middle}$	<		arg$_2$	arg$_3$	>
			[-r]	[+o]	
			|	|	
			SUBJ	OBJ	
			DAT	NOM	

Table 10: Linking Schema for experiencer 'find'

This analysis accounts for the fact that experiencer verbs in Icelandic are often found with the *-st* middle morphology and makes sense of our corpus studies, which show that an increase in dative subjects was mainly associated with middles.

[15] See (Wood, 2015, 231) for argumentation along similar lines. Also see Maling's (1990) example (61), which shows that the dative argument in experiencer middles is not derivable from the corresponding active clause.

5 Conclusion

The overall picture that emerges from our diachronic corpus study suggests that dative case is becoming more systematically associated with lexical semantic meanings and that the overall system is quite complex. On the one hand, datives are becoming more systematically associated with experiencers and goals. On the other hand, the dative is used when the external initiator of an event only initiates an event, but does accompany the object to a result state. Both types of datives feed into valency alternations conditioned by voice. Our corpus study implicates middles as a major factor in the increase of dative subjects over time. We suggested that this is due to experiencers being introduced via applied datives into experiencer/perception predications. In analogy to developments in Latin, we show how such experiencer predicates can arise out of middles for certain verbs and can thus account for the preponderance of middle morphology with experiencer predicates.

While it could be possible that a single Dative Subject Construction has been inherited from Proto-Indo European and now instantiates the various dative subject examples, the data instead point to a complex system in which lexical and event semantic considerations interact with systematic but variable linking possibilities between arguments and grammatical relations, giving rise to several different realizational possibilities and the concomitant introduction of new predicational meanings such as those associated with dative subject experiencer predicates based on verbs of perception.

References

Anderson, Stephen R. 1990. The grammar of Icelandic verbs in *-st*. In J. Maling and A. Zaenen (eds.), *Modern Icelandic Syntax*, pages 235—273, San Diego: Academic Press.

Andrews, Avery D. 1976. The VP Complement Analysis in Modern Icelandic. In *Proceedings of the North East Linguistic Society (NELS)*, volume 6, pages 1–21, reprinted 1990 with minor revisions in *Modern Icelandic Syntax*, 165–185.

Andrews, Avery D. 2001. Non-canonical A/S Marking in Icelandic. In M. Noonan (ed.), *Non-Canonical Marking of Subjects and Objects*, pages 85–111, Amsterdam: John Benjamins.

Arka, I Wayan. 2015. Constructed Middles in Marori: an LFG Analysis. This Volume.

Barðdal, Jóhanna. 2001. *Case in Icelandic — A Synchronic, Diachronic and Comparative Approach*. Ph. D.thesis, Lund University.

Barðdal, Jóhanna. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic*. Amsterdam: John Benjamins.

Barðdal, Jóhanna. 2011. The Rise of Dative Substitution in the History of Icelandic: A Diachronic Construction Grammar Account. *Lingua* 121(1), 60–79.

Barðdal, Jóhanna and Eythórsson, Thórhallur. 2009. The Origin of the Oblique

Subject Construction: An Indo-European Comparison. In V. Bubenik, J. Hewson and S. Rose (eds.), *Grammatical Change in Indo-European Languages*, pages 179–193, Amsterdam: John Benjamins.

Barðdal, Jóhanna, Smitherman, Thomas, Bjarnadóttir, Valgerður, Danesi, Serena, Jenset, Gard B. and McGillivray, Barbara. 2012. Reconstructing Constructional Semantics: The Dative Subject Construction in Old Norse-Icelandic, Latin, Ancient Greek, Old Russian and Old Lithuanian. *Studies in Language* 36(3), 511–547.

Barron, Julia. 2001. Perception and Raising Verbs: Synchronic and Diachronic Relationships. In M. Butt and T. H. King (eds.), *Time Over Matter*, pages 73–103, Stanford: CSLI Publications.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell.

Butt, Miriam, Bögel, Tina, Kotcheva, Kristina, Schätzle, Christin, Rohrdantz, Christian, Sacha, Dominik, Dehe, Nicole and Keim, Daniel. 2014. V1 in Icelandic: A multifactorical visualization of historical data. In *Proceedings of VisLR: Visualization as added value in the development, use and evaluation of Language Resources*, Workshop at the 9th edition of the Language Resources and Evaluation Confernece (LREC 2014), Reykjavik, Iceland.

Butt, Miriam and Deo, Ashwini. 2013. A Historical Perspective on Dative Subjects in Indo-Aryan, Paper presented at the LFG13 Conference.

Dalrymple, Mary. 2001. *Lexical Functional Grammar*. New York: Academic Press.

Deo, Ashwini. 2003. Valency Change and Case Marking: Marathi Dative Experiencers, handout from the Pioneer Workshop on Case, Valency and Transitivity.

Faarlund, Jan Terje. 1994. Old and Middle Scandinavian. In E. König and J. van der Auwera (eds.), *The Germanic Languages*, Chapter 3, pages 38–71, New York: Routledge.

Haspelmath, Martin. 2001. Non-Canonical Marking of Core Arguments in European Languages. In A. Y. Aikhenvald, R.M.W. Dixon and M. Onishi (eds.), *Non-Canonical Marking of Subjects and Objects*, pages 53–83, Amsterdam: John Benjamins.

Haugen, Einar. 1984. *Die skandinavischen Sprachen: Eine Einführung in ihre Geschichte*. Hamburg: Buske.

Hock, Hans Henrich. 1990. Oblique Subjects in Sanskrit. In M. Verma and KP Mohanan (eds.), *Experiencer Subjects in South Asian Languages*, pages 119–139, Stanford: CSLI Publications.

Jónsson, Jóhannes Gísli. 2003. Not so quirky: on subject case in Icelandic. In E. Brandner and H. Zinsmeister (eds.), *New Perspectives on Case and Case Theory*, pages 129–164, Stanford: CSLI Publications.

Jónsson, Jóhannes Gísli. 2013. Dative vs. accusative and the nature of inherent case. In B. Fernández and R. Etxepare (eds.), *Variation in Datives. A Micro-Comparative Perspective*, pages 144–180, Oxford: Oxford University Press.

Kaufmann, Ingrid. 2007. Middle Voice. *Lingua* 117, 1677–1714.

Keim, Daniel, Andrienko, Gennady, Fekete, Jean-Daniel, Górg, Carsten, Kohlham-

mer, Jórn and Melançon, Guy. 2008. Visual Analytics: Definition, Process, and Challenges. In A. Kerren, J. T. Stasko, J.-D. Fekete and C. North (eds.), *Information Visualization*, pages 154–175, Berlin: Springer.

Kemmer, Suzanne. 1993. *The Middle Voice*. Amsterdam: John Benjamins.

Kibort, Anna. 2013. Objects and Lexical Mapping Theory. In M. Butt and T. H. King (eds.), *Proceedings of the LFG13 Conference*, CSLI Publications.

Kibort, Anna. 2014. Mapping out a construction inventory with (Lexical) Mapping Theory. In M. Butt and T. H. King (eds.), *Proceedings of the LFG14 Conference*, pages 262–282, CSLI Publications.

Levin, Beth. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: Chicago University Press.

Maling, Joan. 1990. The hierarchical assignment of grammatical cases, handout from talk at Stofnun Sigurðar Nordals.

Maling, Joan. 2002. Það rignir þágufalli á Íslandi. Verbs with Dative Objects in Icelandic. *Íslenskt mál* 24, 31–105.

Marcus, Mitchell P., Marcinkiewicz, Mary Ann and Santorini, Beatrice. 1993. Building a large annotated corpus of English: the Penn Treebank. *Computational Linguistics* 19(2), 313–330.

Schätzle, Christin, Sacha, Domnik and Butt, Miriam. 2014. Diachronic Visualization of Dative Subjects in Icelandic, poster presentation at the Workshop on Big Data Visual Computing, *44th Annual Meeting of the Gesellschaft für Informatik*.

Sigurðsson, Halldór Ármann. 1989. *Verbal Syntax and Case in Icelandic: A Comparative GB Approach*. Ph. D.thesis, Lund University.

Smith, Henry. 1994. "Dative Sickness" in Germanic. *Natural Language and Linguistic Theory* 12(4), 675–736.

Smith, Henry. 1996. *Restrictiveness in Case Theory*. Cambridge: Cambridge University Press.

Svenonius, Peter. 2002. Icelandic case and the structure of events. *Journal of Comparative Germanic Linguistics* 5, 197–225.

Svenonius, Peter. 2006. Case Alternations and the Icelandic Passive and Middle. In S. Manninen, D. Nelson, K. Hiietam, E. Kaiser and V. Vihman (eds.), *Passives and Impersonals in European Languages*, Amsterdam: John Benjamins.

Thráinsson, Höskuldur. 1994. Icelandic. In E. König and J. van der Auwera (eds.), *The Germanic Languages*, Chapter 6, pages 142–188, London: Routledge.

Wallenberg, Joel C., Ingason, Anton Karl, Sigurðsson, Einar Freyr and Rögnvaldsson, Eiríkur. 2011. Icelandic Parced Historical Corpus (IcePaHC).

Wood, Jim. 2015. *Icelandic Morphosyntax and Argument Structure*. New York: Springer.

Zaenen, Annie and Maling, Joan. 1984. Unaccusative, passive, and quirky case. In M. Cobler, S. MacKaye and M. T. Wescoat (eds.), *Proceedings of the Third West Coast Conference on Formal Linguistics (WCCFL)*, pages 317–329.

Zaenen, Annie, Maling, Joan and Thráinsson, Höskuldur. 1985. Case and grammatical functions: the Icelandic passive. *Natural Language and Linguistic Theory* 3(4), 441–483.

INDIVIDUATING LEXEMES IN LFG

Andrew Spencer
University of Essex

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King (Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

Lexicalist models of morphosyntax need to have a way of individuating lexemes. I adopt Spencer's (2013) Generalized Paradigm Function Morphology (GPFM) model of lexical relatedness, in which lexemes/lexical entries are individuated by means of an arbitrary index, distinct from phonology, syntax or semantics. I sketch an LFG interpretation of the GPFM model of argument structure which crucially appeals to semantic function (SF) roles to specify morphosyntactic category. I focus on transpositions, deploying the notion of complex SF role, (which can be used to label corresponding c-structure nodes). I distinguish true transpositions and meaning-bearing transpositions from transpositional lexemes, distinct lexemes which do not involve an added semantic predicate. I illustrate the latter with deverbal adjectives such as English *interesting*. The analysis crucially appeals to GPFM's LI attribute. I show that the LI attribute is virtually coextensive with LFG's PRED attribute. GPFM and LFG are therefore compatible.

1 Introduction

Linguistic theories, especially lexicalist models, need an explicit account of the lexicon, that is, an explicit characterization of what information goes into lexical representations and how lexical entries are related to each other ('lexical relatedness').* This has not been a major focus of research in LFG but it is a prerequisite for a study of morphology/word structure and especially for a study of the relation between word structure and phrase/sentence structure. A case in point is the problem of 'mixed categories'. The most widespread of these are the 'transpositions', illustrated by (productive) action nominalizations (*his writing the letter so quickly*) and deverbal participles (*the girl quietly reading a book in the corner*). These typically have the 'external syntax' of a derived morphosyntactic category while retaining a good deal of the syntax (and semantics) of the base category ('internal syntax').

I summarize a model of lexical representation based on Stump's (2001) Paradigm Function Morphology (PFM), and the model of Generalized Paradigm Function Morphology (GPFM) proposed in Spencer (2013), which extends the model of lexical relatedness implicit in PFM to all types of (paradigmatic, regular) relatedness. An important set of questions for the GPFM model centres around the issue of how to distinguish one lexeme (or lexical entry) from another. This is important in lexeme-based models such as PFM, GPFM and, in fact, any model which effectively draws an architectural distinction between inflection (intra-lexemic relatedness) and derivation (inter-lexemic relatedness). However, it is well-known that there is no principled way to distinguish inflection from derivation (Spencer 2013). This means that there is no automatic way of deciding when two word forms are forms of the same lexeme or of different lexemes.

The GPFM model makes considerable appeal to a notion of Lexemic Index (LI), an arbitrary integer serving to individuate lexemes. Other models adopt sig-

*I am grateful to Yoko Sugioka, to an external reviewer and the editors for very helpful discussion.

nificantly different ways of individuating lexemes. I argue that the GPFM Lexemic Index essentially fulfills the same function as LFG's PRED attribute, as it is used in the current architecture. This means that it should be relatively more easy to integrate the GPFM model into LFG architecture than models of morphology (such as those discussed below in Section 4) which adopt different strategies for individuating lexical entries.

I make explicit a number of assumptions that remained implicit in Spencer (2013), specifically concerning the typology of lexical relatedness, but also regarding the details of how transpositions are to be analysed in the model. I then consider the notion of 'transpositional lexeme', briefly introduced in Spencer (2013) extending the empirical base which motivates that notion.

2 Lexical representations and lexical relatedness

2.1 Lexical representations

The GPFM model of lexical relatedness (Spencer, 2005, 2013) assumes that a lexeme is a simple four-dimensional vector, ⟨FORM, SYN(TAX), SEM(ANTICS) L(EXEMIC) I(NDEX) ⟩, which can be represented schematically as in (1):

(1) Lexical entry as four-dimensional vector

FORM:	representation of the (inflected) word form
SYN:	grammatical function array argument structure collocational/selectional restrictions other syntactic properties
SEM:	e.g. Lexical-Conceptual Structure
LI:	a unique integer (LI)

Here, I assume that FORM means:

- a specification of the 'root' of the lexeme and any non-predictable stem forms (this effectively subsumes phonological information)
- a specification of any non-predictable inflected word forms
- a pointer to an (inflectional) morphology engine (cf. PFM)

SYN means:

- GF structure: a specification of the array of grammatical functions associated with a predicate's entry
- argument structure: a list of the predicate's (external and internal) arguments (possibly with thematic role labels)
- a specification of other types of selection, for instance, whether a verb takes an infinitival complement, an *ing*-complement, or both

SEM means: Some appropriate lexical semantic representation. In Spencer (2013) I assume some paired-down version of Jackendoff's Lexical Conceptual Structure (Jackendoff, 1990; Levin and Rappaport Hovav, 2005), including Jackendoff's ontological categories Thing, Event, Property. Note: this type of conceptual/lexical representation is to be distinguished from 'denotational', 'model theoretic', 'type theoretic' semantic representation (see Bach and Chao, 2009, 2012 on this distinction). In practice, this means that two SEM representations might be 'equivalent' when the corresponding typed representations are distinct, since the latter change with morphosyntactic context. For instance, a noun such as *cat* in *catfood* is treated here as semantically identical to the *cat* of *a black cat*, even though in the noun-noun compound *cat* is an attributive modifier and so it will presumably be of type ⟨⟨e,t⟩,⟨e,t⟩⟩, not ⟨e,t⟩.

The LEXEMIC INDEX (LI) corresponds in function to what Sag (2012) refers to as the Lexical Identifier (LID). However, in GPFM it is a unique arbitrary integer identifying each distinct lexeme, somewhat in the manner of a key in a database.

A schematic example of a lexical entry is given in (2).

(2) Schematic/simplified entry for 'write'

FORM: /raɪt, roʊt, rɪtn̩/

SYN: ⟨SUBJ, OBJ⟩
⟨Ext-arg, Int-arg, e⟩
Int-arg = linguistic/musical text

SEM: λx,y[$_{\text{Event}}$ **write′**(x,y)]

LEXEMIC INDEX: 59 (or more perspicuously, WRITE)

The precise definition of 'argument structure representation' is not relevant to this discussion, and so I will just represent arguments as ⟨ARG1, ARG2, …⟩. The crucial point is that argument structure includes a set of 'semantic function roles' (SF roles): the 'R' role for nouns (Williams, 1981), the event role, 'E', for verbs (as found in neo-Davidsonian approaches to lexical semantics, and also the approaches such as that of Bierwisch, 1983, Wunderlich, 1997). For adjectives I follow typological tradition (e.g. Croft, 2003, see also Spencer, 1999, 2005) in assuming that the primary function of adjectives as that of attributive modification (of a noun). Therefore, the SF role of an adjective is 'A*', where the '*' is a notational device to indicate that the adjective's SF role is coindexed with 'R', the SF role of the noun it modifies. For instance, a representation such as *Swedish* ⟨A*$_x$⟨x⟩⟩ *man* ⟨R*$_x$⟩ will correspond roughly to the representation λx(**Swedish**(x) ∧ **man**(x)) (see Spencer, 2013, 33f for further discussion and motivation).

The SF roles largely, if not entirely, subsume the functions of c-structure labels, N, V, A. The rationale for adopting this view of SF roles is:

(i) tradition: the 'R' and 'E' roles are already widely deployed in the literature (e.g. Lieber, 2004
(ii) the problem of mixed morpholexical categories (especially transpositions).

Spencer (1999) proposes representing transpositional categories such as action nominals, relational adjectives and deverbal participles with complex or composite SF roles. Thus, an action nominal derived from a verb with a-structure ⟨E⟨…⟩⟩ is the name of an event predicate and has the composite SF role ⟨R⟨E⟨…⟩⟩⟩, while a participle is the attributive modifier representation of that predicate: ⟨A*⟨E⟨…⟩⟩⟩.

We often find that the inflectional paradigm of a lexeme includes periphrases, that is, combinations of a form of the main lexeme together with inflected forms of one or more functional (auxiliary, ancillary) lexemes. Moreover, some lexemes have a basic form which itself consists of more than one lexeme. A clear example of this would be lexicalized but nonetheless transparent compounds, such as *catfood, lightpen, men's room*, … as well as productive noun incorporation in many languages. Other cases include many light verb constructions (*render assistance to*), some verb-particle constructions (*pull apart*), and so on. This means that the paradigm function defining the word forms of a lexeme will sometimes define multiword expressions (better, multilexeme expressions, MLEs), and each of the component lexemes (main verb and auxiliary verb, for instance) will have its own lexical entry. A lexical entry for such a multiword lexeme will therefore itself be defined in terms of a *set* of lexical entries. Now, in GPFM, the basic lexical entry of a lexeme is itself the result of applying the paradigm function to a pairing of Lexemic Index and a null feature set, effectively capturing the idea that the lexical entry is the starting point for the definition of an inflectional paradigm, and the starting point for all the derivational morphology defined with that lexeme as its base. This means that a multiword lexeme will be similar to a periphrastic inflected word form, but one whose parts are the realization solely of the lexical meaning, not of any functional/inflectional features. What this means is that we can conflate the terms 'lexeme' and 'lexical entry', provided we bear in mind that lexemes can be complex and themselves consist of lexemes, and that those MLEs can be the result of productive, paradigmatic lexical processes, so that the complex lexemes themselves are not (necessarily) 'listemes'.[1]

2.2 Lexical relatedness

There is a long tradition of discussing the way that word meanings are related to each other and this is the basis of much lexicography, thesaurus construction and lexicology generally (Cruse, 1986), but words can be related in a variety of ways, depending how we understand the notion 'word' and 'related'.

First, we must distinguish lexemes from (inflected) word forms of a lexeme. For instance, the word forms {*write*, *writes*, *wrote*, *writing*, *written*} are all related

[1] But see Stump, forthcoming, for a dissenting view.

to a lexeme, WRITE, by virtue of realizing the FORM attribute of that lexeme's lexical entry. These forms are also all related to each other in purely formal terms, either by simple affixation (*write, writes, writing*), irregular (unpredictable) ablaut (*write, wrote*), or both (*write, written*). By the same token, the forms *writes, written* or *writing, wrote* are lexically related, albeit indirectly (via the root or stem of the lexeme). FORM-based relatedness cross-cuts syntactic and semantic relatedness and clearly is independent of lexemic identity: thus, even if *re-write* is a form of a distinct lexeme from *write* the two are still related by FORM (as well as by SEMANTICS). Some words are related by FORM without enjoying any other relationship. Thus, the various readings of the homonymous lexemes DRAW (as in *draw a picture* vs. *draw blood, conclusions, breath*) share exactly the same array of inflected forms. Similarly, the verb UNDERWRITE shares all the forms of WRITE together with the meaningless prefix *under-* even though they share nothing else.

At the level of syntax two words may be related in a variety of ways. They may be argument structure variants of each other (*give Fido a bone* ∼ *give a bone to Fido*, they may be semantically related in a way that has syntactic repercussions (*The cakes burned* ∼ *Alfred burned the cakes*), they may be related constructionally by addition of an extra argument (*They baked her a cake, She whistled her way out of the room*. For a given language and for a given pair of alternants it remains an open question whether we should regard the alternants as variants of the same lexeme or not.

The examples just cited also illustrate ways in which words may be semantically related. In canonical derivation we see the SEM representation of some base lexeme enriched by addition of a semantic predicate, as in WRITER, 'person/thing realizing the SUBJECT argument of the verb WRITE'. In some cases the meaning relation is not expressed formally, in which case we may speak of zero derivation or of coercion. Much of the literature on polysemy, including that inspired by the notion of the Generative Lexicon (Pustejovsky, 1995) revolves around the question of how to define relatedness at the semantic level. Where we have what Apresjan (1974) calls 'systematic polysemy' (*reguljarnaja mnogoznačnost'*) we are, presumably, dealing with distinct interpretations for a single lexeme. This contrasts with 'accidental polysemy' as in *branch (of a tree)* vs. *branch (of a bank)*, which presumably is effectively homonymy between two distinct (though formally related) lexemes, but the problem lies then in distinguishing systematic from accidental polysemy.

A preliminary summary of lexical relatedness types is provided in (3).

(3) Two words $£_1$, $£_2$ are lexically related iff
- $FORM(£_1) = \Phi(FORM(£_2))$, where Φ is some morphophonological relation); and/or
- $SYN(£_1) = \Sigma(SYN(£_2))$, where Σ is some operation over syntactic representations; and/or
- $SEM(£_1) = M(SEM(£_2))$, where M is some operation over semantic representations

The Φ operation defines systematic allomorphy (that is, allomorphy that is not suppletive). The Σ operation will include argument structure alternations and possibly other types of systematic syntactic operation. The M operation defines (presumably monotonic increasing, Koontz-Garboden, 2007) alterations which we can think of informally as 'adding a semantic predicate'. Non-trivial application of M will therefore mean that SEM($£_1$) ⊂ SEM($£_2$), where '⊂' means that SEM($£_2$) is an 'extension' of SEM($£_1$). In the most general case, this means that the set of entailments associated with SEM($£_2$) properly includes the set of entailments associated with SEM($£_1$). Two lexical items will also be lexically related if LI($£_1$) = Λ(LI($£_2$)), for some function, Λ, but it will not be necessary to include this in the definition, because any such pair will also be related in terms of FORM and/or SYN and/or SEM representations according to (3). In the special cases where (specific applications of) Φ, Σ, M, Λ instantiate the identity function we have straightforward cases of relatedness.

We may now ask what restrictions there might be on the way that two arbitrary lexical items might be related. Spencer (2013, 139) proposes the maximally simple answer to this question, (4):

(4) *Principle of Representational Independence (PRI)*:
Each of the four principal attributes or dimensions of a lexical representation can define lexical relatedness independently of the other three.

A clear example of the application of the PRI is found with a very common but relatively understudied type of lexical relatedness, the transposition. Here, the morphological and syntactic category is changed but the lexical content (SEM) remains the same. A true transposition can be shown to be a form of the base lexeme, which means that shares its LI with the base. The most studied transposition is the action nominal (Koptjevskaja-Tamm, 1993; Vendler, 1968; Zucchi, 1993; Alexiadou, 2010a,b and for LFG Nikitina, 2008). Spencer (2005) shows that we can find productive instances of all of the six logically possible ways of transposing the major lexical categories, N, V, A, into each other.

Spencer (2013) follows a number of authors in coding grammaticalized, productive transpositions in terms of a category of REPRESENTATION (REPR, cf. the Russian term *representacija*): an action nominal is the nominal representation of a verb lexeme. Despite the change in category it remains a form of a verb and does not change the lexemic status of that verb: *writing* is a form of the verb lexeme WRITE, despite being categorially a noun (in part), or an adjective (in part). It is thus a feature of a transposition that it shares the Lexemic Index of its base. Thus, in a (canonical) transposition, FORM($£_1$) = Φ(FORM($£_2$)), SYN($£_1$) = Σ(SYN($£_2$)) for non-trivial Φ, Σ, but SEM($£_1$) = SYN($£_2$) and, crucially, LI($£_1$) = LI($£_2$).

3 Lexical representations and the Lexemic Index in LFG

The transpositions (which Kuriłowicz, 1966 referred to as 'dérivation syntaxique') raise significant questions for the way that lexical representations interact with syntactic representations. The architecture of LFG is particularly well suited to exploring these interactions because of the way that it factorizes parallel aspects of syntactic structure mirroring the factorization of lexical structure in GPFM (the parallel between LFG and GPFM is not entirely accidental, of course). In Spencer (2013) I sketch a 'proof of concept' of the way that the GPFM representations might be coded using LFG feature structures and this is deployed for the analysis of participles in Spencer (this volume).

The GPFM architecture relies heavily on being able to distinguish forms of a single lexeme from (related) forms of distinct lexemes. We will see later how lexemic individuation can be used to characterize rather subtle distinctions in word behaviour across languages. The LI is a useful formal (or at least, notational) device for recording when we regard two forms as belonging to the same lexeme or not.

In LFG lexemes were not originally identified as linguistic objects: and the main interface between lexicon and syntax was mediated through word forms, not lexemes. The nearest thing to a formal specification of lexemehood was the PRED value, which, however, united a lexeme's name, its grammatical functions (where appropriate) and some (implicit) characterization of its meaning. In contemporary LFG the semantic properties of the lexeme/lexical entry are defined in terms of the semantic projection σ (Andrews, 2008, Asudeh, 2012). In practice, this leaves only the kinds of PRED functions sketched in (5).

(5) Functions of the PRED value:

For nouns	[PRED 'man'], [PRED 'Named: "Harriet"']
For adjectives:	[PRED 'Swedish⟨SUBJ⟩']
For prepositions:	[PRED 'between⟨SUBJ, OBJ⟩']
For verbs:	[PRED 'give⟨SUBJ, OBJ1, OBJ2⟩']

On this view, the PRED value therefore provides (exactly) two sorts of information:

(i) A specification of the grammatical function array associated with the predicate: GF-ARRAY = ⟨SUBJ, OBJ1, OBJ2⟩
(ii) A unique 'name' for the lexeme: 'man', 'Harriet', 'Swedish', 'between', 'give' (cf. PRED FN)

We can therefore split the PRED attribute into two attributes, GF-ARRAY and 'unique name', as shown in (6). The 'unique name' is, of course, the LI. The GF-ARRAY is a specification of the 'construction type' that the lexeme belongs to (intransitive verb, transitive adjective and so on).

(6) Factorization of PRED attribute

PRED ‘give⟨SUBJ, OBJ1, OBJ2⟩’ ⇒
LI: GIVE
GF-ARRAY: ⟨SUBJ, OBJ1, OBJ2⟩

PRED ‘seem⟨XCOMP⟩ SUBJ’ ⇒
LI: SEEM
GF-ARRAY: ⟨XCOMP⟩ SUBJ

Note: A true transpositions does not have a distinct PRED FN value (LI) from the base lexeme, even if its a-structure and GF-ARRAY are different from that of the base.

4 Individuating lexemes: various views

In this section I consider three contrasting proposals for individuating lexemes. In the first, the lexeme is identified in terms solely of its root. In the second, the lexeme’s meaning is the principal marker. I argue that both these approaches are flawed. The third alternative is conceptually the simplest: the LI is an arbitrary label (for instance, a unique integer) with no other properties, as proposed in Stump (2002) and particularly in Spencer (2013).

4.1 The simplest set of assumptions

In the first model, all inflection and derivation is realized by separate ‘functional heads’ (‘morphemes’) each with its own lexical entry. All morphology is compounding, and compounding is an operation in syntax (‘merge’). This is the essence of standard Distributed Morphology (DM) (Marantz, 1997).

In terms of the GPFM model sketched above, we can say that, in DM, each distinct FORM representation corresponds uniquely to a single SYN, SEM representation.[2] An entailment of this model is that no lexical items whatsoever, in any language whatsoever, can show (root) suppletion (see Borer, 2013, 398-99 for a very explicit statement). This appears to be untrue (Corbett, 2007, Harley, 2014, Spencer, to appear).

It is very difficult to see how such a model could account for the more articulated aspects of lexical structure discussed here without smuggling into the model some equivalent to the LI, so I will ignore such models from now on.

[2] In fact, it is generally assumed that this is true only for lexical roots, not for functional heads. However, it is very unclear how that lexical∼functional distinction is supposed to be drawn in this model.

4.2 Second possibility: take SEM value/representation as its index

Sag (2012) deploys an attribute 'Lexical Identifier' (LID) in his model of lexical structure. The LID is an attribute of the HEAD|SYN attribute. It is, however, tagged with the semantic representation in such a way that two lexical entries with the same meaning (including synonyms) will have the same LID. This in itself is only a problem if we believe that there are true synonyms. However, it does mean that the LID has to be tied to rather subtle aspects of CONTENT and usage (including collocations) if it is to be a coherent notion. There are, however, other problems with Sag's conception of the LID.[3]

The first problem is that there is morphology which changes a lexeme's meaning but without changing its LID. There are regular argument structure alternations such as productive causative forms of verbs, and below I discuss the case of meaning-bearing noun-to-adjective transpositions in Selkup. There is also a rather subtle technical problem with Sag's LID: the SBCG model inherits from standard HPSG a distinction between linguistic objects and their descriptions. For Sag a lexical entry is a description of an object, a lexeme, which is a member of the types of linguistic expression. However, Sag encounters difficulties in providing an illustration of a lexeme. This is because HPSG is essentially defined in terms of word forms. Sag therefore has to take an arbitrarily chosen inflected form of a lexeme in order to illustrate it as a linguistic object (the preterite form *laughed* to illustrate the lexeme LAUGH, Sag 2012, 99–101). But that means that Sag is talking about word forms, not the lexemes of which those words are forms. But it is now not possible to see how the LID is individuating lexemes as opposed to word forms of lexemes. I can see no way around this conceptual difficulty.

Bonami's (2015) analysis of periphrasis deploys the LID to express the idea that an auxiliary verb is part of the realization of an inflectional property in a (true) periphrasis. For instance, in the Latin perfect passive periphrasis *amatus est* '(he) is/has been loved' we would say that SUM ('be') had INFL attribute with LID SUM, but also has its HEAD|LID feature tagged with the HEAD|LID of AMO ('love'). This expresses the intuition that the auxiliary lexeme LID SUM is used to realize part of the inflectional paradigm of a lexical verb such as the lexeme 'love', LID AMO. However, as Bonami points out, this device will not work if the LID is tied to the SEM representation, because, in LFG terms, it would give rise to a clash of (old-style) PRED values.

4.3 LI = arbitrary stipulated label (e.g. integer)

To avoid potential confusion I shall from now on refer to the arbitrary indexing marker for a lexeme as the Lexemic Index (LI), following Spencer (2013). It is the LI that serves to record particular decisions about how lexemes are individuated. The LI on this account is an attribute of a lexeme, distinct from the FORM, SYN,

[3] Expletive elements such as *it, there* presumably require LIs (as a reviewer reminds me) but it is difficult to see how that can be achieved if the LI is tied to a semantic representation.

SEM attributes.[4]

5 Transpositions and transpositional lexemes

Assuming the feature system sketched in Spencer (2013), in which REPR defines (canonical) transpositions, along with the notational conventions used there, a participle is defined as the result of applying the generalized paradigm function to the pairing $\langle \mathcal{V}, \rho \rangle$, where $\mathcal{V}$ is the LI of a verb and ρ = {[REPR[TRANS2A[V2A]]], …}. This function defines a new set of word forms and crucially redefines the SF role from $\langle E\langle x,\ldots\rangle\rangle$ to $\langle A^*_x,\langle E\langle x,\ldots\rangle\rangle\rangle$. Being an adjective the participle now has a set of inflected forms which are entirely absent in finite verbs forms, namely, the agreements in features such as gender, number, case (and definiteness, in the case of Baltic languages; see Spencer, this volume). This means that the inflectional morphology must treat them as though they were distinct lexical entries (even though they remain in an important sense 'forms of the base verb lexeme'). Spencer (2013) argues that the morpholexical categories which are defined by the SF roles are associated by default with a morpholexical signature. This specifies precisely which properties a word is inflected for. Since all the transposed participial forms are now categorially adjectives (as well as, to some extent, still verbs) they are subject to that default specification. It is a matter of language-particular stipulation precisely how many of the verbal properties are retained and in what form.

Adjectives in a language such as Latin, Sanskrit or Lithuanian are subject to a default specification under which their SYN attribute bears an agreement attribute, AGR (Haug and Nikitina, 2012, Spencer, this volume). This default specification will now apply to the participle's SYN attribute, meaning that appropriate agreement relations can be defined in the syntax. By a further default specification, the syntactic agreement AGR attribute is copied in the FORM attribute, as a morphological AGR attribute. This ensures that the inflectional morphology will provide the actual inflected forms required by the syntactic representations. In the case of languages such as Sanskrit and Lithuanian the FORM|AGR attribute will be further specified with information specifying which inflectional and accentual class the participle belongs to (just as in the case of ordinary adjectives).

The discussion of participles in Spencer (2013) is somewhat sketchy and so I will here flesh out some of the details of how such transpositions are to be handled (the general principles will apply to all six types of true transposition). Specifically, I clarify an important detail concerning the way that inflected forms of transpositions are defined.

In Stump's (2001) PFM model the paradigm function for a lexeme defines a single set of inflectional properties for that class of lexemes. In Spencer (2013) I

[4] In GPFM, the functions which define the FORM, SYN and SEM values of a lexeme or word form are defined in terms of the LI. Thus, if we assume a LI CAT, the form /kat/, the SYN value [REF[ARG1[ ]]], and the SEM value λx.**cat**(x) are related as values of FORM$\langle$CAT,{Ø}$\rangle$, SYN$\langle$CAT,{Ø}$\rangle$, SEM$\langle$CAT,{Ø}$\rangle$ (where '{Ø}' is the null feature set). Details can be found in Spencer (2013), but they are not directly relevant to the point being made in this paper.

appeal to a morpholexical signature, a declaration of precisely those properties for which a given class inflects, notated as value, MORSIG, of the FORM attribute in the lexical entry. By the *Principle of Inflectional Specifiability* (Spencer 2013, 199) a lexeme has to be furnished with such a signature before it can be inflected. But this means that transposition will in general bear the wrong morpholexical signature. For instance, a deverbal participle will bear the signature of a verb, not that of an adjective. If we treated the participle as though it were purely derivational we could simply overwrite the verb's morpholexical signature and replaced it with that of an adjective. However, then we would then lose the fact that the participle retains crucial verb properties. In this respect the participle is a prototypical instance of a paradigmatically mixed category and this has to be reflected in its morpholexical signature.

Unfortunately, very little of cross-linguistic validity can be said about paradigmatic categorial mixing. Malchukov (2004, 2006) has very interesting observations on inflectional property mixing in action nominals. He asks to what extent a verb loses its verbal properties and to what extent it gains nominal properties. We can ask the same question about participles (cf. also Lehmann, 1984) but that work has yet to be done. Although there are strong tendencies apparent, there is, it would appear, no way to predict precisely which verb properties will be lost and which adjective properties will be acquited. The precise content of the MORSIG therefore has to be stipulated on a language-by-language basis.

There are two aspects to defining inflections: (i) what properties does the lexeme inflect for? and (i) what are the actual inflected word forms? The word forms will depend on the inflectional and morphological class the word belongs to. In Lithuanian there is a clearly defined morphological class of Adjective (with various declensional and accentual subclasses). The MORSIG attribute therefore includes an attribute MORCLASS which defines the word's [MCLASS: Adjective] and its inflectional (declensional) and accentual) class, INFLCLASS, ACCENTCLASS. At the same time the MORSIG bears an attribute INFL which specifies the agreement properties of an adjective. This is shown in Figure 1, for the adjective *báltas* 'white'.

A verb will have a MORSIG attribute which specifies a MORCLASS|MCAT value Verb, and INFL properties of subject agreement and tense-aspect-mood-voice inflection, as seen in Figure 2, for the verb *dìrbti* 'work' (with the most common feature values included for illustration).

The representation of the present active participle is shown in Figure 3 from the verb *dìrbti*.

The morphology of Lithuanian defines a participial stem form (or forms) for each class of verbs and these stems are then inflected like adjectives. Thus, to define the present active participle (PPA) of a verb $\mathcal{V}$ we need the Generalized Paradigm Function (GPF) shown in (7). The f_{form} component of the GPF defines the FORM template representation shown in Figure 3. The f_{syn} component defines the compound A-STR representation. The $f_{sem,\ li}$ components induce no change, of course, and are omitted.

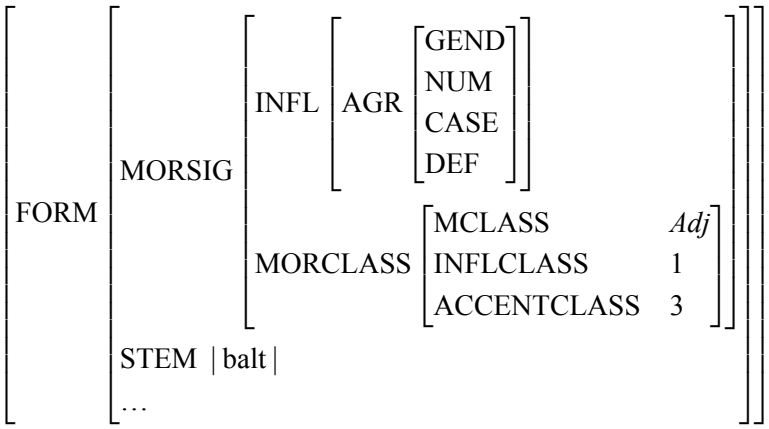


Figure 1: FORM attribute from lexical entry for Lithuanian adjective *báltas* 'white'

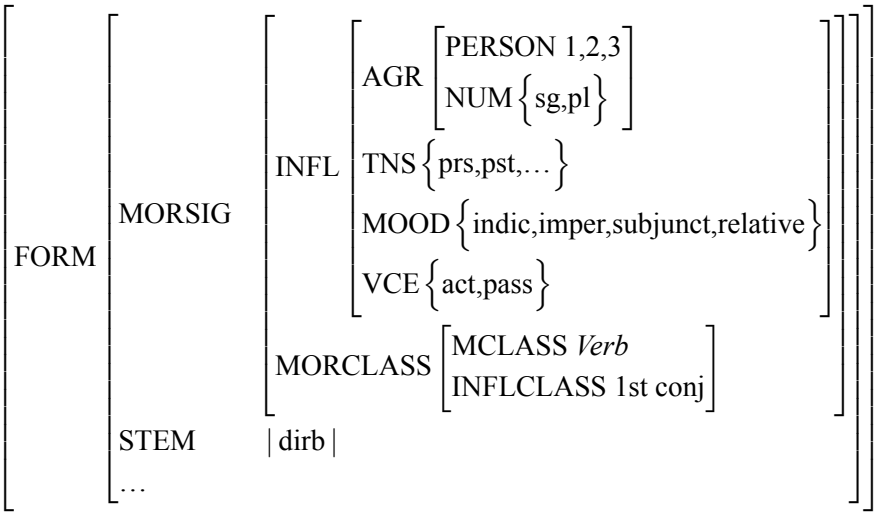


Figure 2: FORM attribute from lexical entry of Lithuanian verb *dìrbti* 'work'

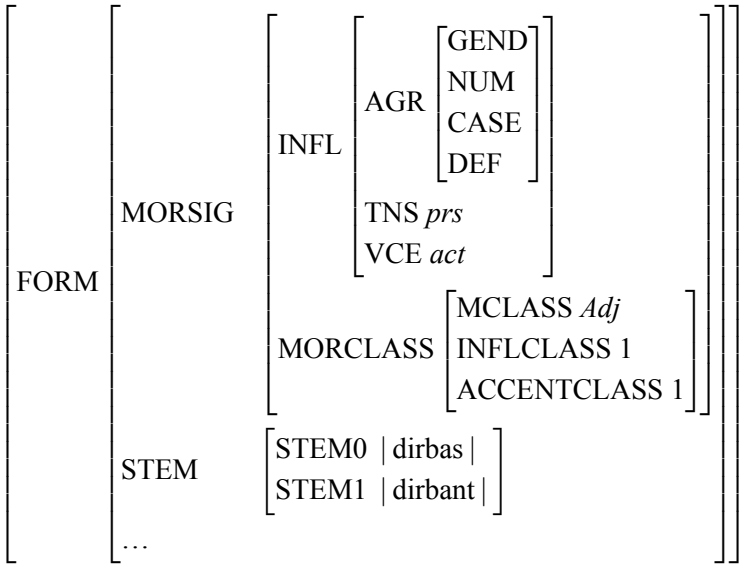


Figure 3: FORM attribute for present active participle form of Lithuanian *dìrbti*, *dirbq̃s* 'working'

(7) $gpf(\langle \mathcal{V}, \{[REPR:V2A], TNS:PRS, VCE:ACT\}\rangle) =$
$f_{form}(\langle \mathcal{V}, \{[REPR:V2A], TNS:PRS, VCE:ACT\}\rangle) \wedge$
$f_{syn}(\langle \mathcal{V}, \{[REPR:V2A], TNS:PRS, VCE:ACT\}\rangle)$

The f_{syn} function essentially adds the A-MOD semantic function role to the EV role in the verb's argument structure. The operation of the f_{form} function is more complicated. First, the function (re)defines the MORSIG attribute of the participle. Some of the verb features are copied to the participle's MORSIG (voice, tense) others are overwritten (all the MORCLASS and AGR features). The MOOD attribute is absent from the participle's MORSIG. The representation in Figure 3 shows the participle's stem set, not the fully realized paradigm. The STEM attribute houses the Stem Formation information needed to define the final forms of the participle, as described in detail by Stump (2001) for the very similar system of Sanskrit participles.

True transpositions have no effect on the SEM or LI attributes of the base lexeme. This is particularly clear in the case of the pure relational adjectives of the Samoyedic language Selkup. These are noun-to-adjective (N2A) transpositions (Spencer, 2013, Chapter 10). In Selkup relational adjectives are derived from nouns marked for possessor agreement inflection, that is, clearly inflected forms of the noun, yet they still change the word's morpholexical class from noun to adjective. Moreover, the base noun is still 'visible' to the syntax, in that it can still be modified by an attributive modifier (adjective), even though it bears the relational adjective suffix and is morphosyntactically now an adjective. This is shown schematically in (8), where (8a) is the morphological structure, while (8b) is the structure implied

by the attribute-modifier syntax of the construction.

(8) a. *old* *canoe*-RELADJ *oar*
b. [[*old* *canoe*]-RELADJ] *oar*
'oar from the old canoe'

Examples such as (8) clearly show that the relational adjective form is a 'mixed category', behaving like a noun with respect to inflection and modification by attribute, but itself functioning as an adjectival attributive modifier.

Selkup has two additional types of N2A transposition, a locative relational adjective ('situated in/at my canoe'), and a similitudinal relational adjective ('similar to my canoe'). These are also derived from possessor marked nouns, so they are also true transpositions, but they involve addition of a semantic predicate, hence, they are not canonical transpositions (canonically, it is only derivational morphology which alters the SEM attribute of an entry). The existence of such transpositions demonstrates (again) that lexemes can't be individuated solely in terms of semantic differentiation: the similitudinal/locative relational adjectives are semantically enriched compared to the base noun lexeme but they are still forms (adjectival representations) of that noun lexeme and not distinct lexemes.

Relational adjectives can also be found in languages such as English, Russian, and many others. However, unlike the relational adjectives of Selkup, these largely lack the original noun properties either inflectional or syntactically. In fact, they behave like distinct lexemes and respect Lexical Integrity. This means that their base noun is not visible to syntactic modification, for instance. Thus, the relational adjective *prepositional*, does not allow its base noun, *preposition*, to be modified by an adjective such as *spatial* (9a), though the noun *preposition* can be modified by *spatial* in the compound noun *preposition phrase*, (9b).

(9) a. *preposition* ⇒ *prepositional* ⇒ *prepositional phrase*
b. *preposition* ⇒ *spatial preposition*
c. *[[*spatial preposition*]*al phrase*
= 'spatial phrase headed by a preposition'
≠ 'phrase headed by a spatial preposition'
d. *spatial preposition* ⇒ *spatial preposition phrase*]
= 'spatial phrase headed by preposition' *spatial* [*preposition phrase*]

Example (9c) can only mean 'spatial phrase headed by a preposition'.

What this suggests is that the English relational adjective *prepositional* and all those like it are now lexemes in their own right, not transposed forms of nouns (after all). And yet they are semantically identical to their base nouns. For this reason the expression *prepositional phrase* is entirely synonymous with the noun-noun compound preposition phrase. Spencer (2103) labels this type of derived lexeme a

'transpositional lexeme'.[5]

The existence of transpositional lexemes is unexpected given standard assumptions about inflection and derivation, because derivational morphology is supposed to create new lexemes with enrichment of the semantic representation, yet the transpositional lexeme has the same SEM respresentation as its base. However, transpositional lexemes are exprected if we accept the Principle of Representational Independence, and, indeed, we would have to invoke some additional principle to exclude them.

It is reasonable to ask whether other types of transpositional relation give rise to transpositional lexemes. Arguably, there is a host of such examples in English. Obvious examples are provided by action nominals. If *shooting* in *Lee Harvey Oswald's shooting Kennedy (was appalling)* is a V2N transposition, then *shooting(s)* in *There has been a shooting/have been a lot of shootings (in Dallas)* is presumably a distinct lexeme, but one whose meaning, 'act(s) of shooting', is non-distinct from that of the verb.

Spencer (2013) discusses the property nominalization (A2N transposition) found in many languages. While in some languages such nominalizations may be true transpositions,[6] showing the same apparent violations of Lexical Integrity found with Selkup relational adjectives, in a language like English we appear to be dealing with a transpositional lexeme. Again, the lexical semantics of a nominal such as *happiness, popularity, ...* may be the same as that of the base adjective but that adjective can no longer be modified (*very popular* vs **very popularity*), nor can inflected forms be nominalized (*happiest* vs **happiestness, *most happiness*).

English has (true) adjectives derived from participles, both active: *interesting, boring, charming, flattering, challenging, ...*, and passive: *broken, disappointed, wounded, enlightened, interested, bored, flattered, challenged, ...* These adjectives sometimes acquire entirely idiosyncratic meanings, but where the basic verb denotes a psychological predicate their basic meaning can be directly related to that of the verb they are (etymologically) derived from, as indicated in (10).

(10) a. *This book is very interesting/This is a very interesting book* $\approxeq$ *this book interests me a great deal*

b. *Her reaction was in some ways surprising/That was an in some ways surprising reaction* $\approxeq$ *Her reaction surprised us in some ways*

The example illustrated in (10a) raises the question of the role of a degree modifier such as *very* in such constructions. At first sight, *interesting* appears to be an ordinary gradable, like *tall*, because it is modified by *very*. Degree modifiers such as *very* cannot modify genuine participles: **the book very interesting me (is about*

[5]Of course, we find considerable variation here. Very often the relational adjective construction will have distinct nuances from compounds or prepositional phrase structures: *mathematics course* $\neq$ *course of mathematics* $\neq$ *mathematical course*. In other cases, one or other expression will be simply unavailable: *religious history* $\sim$ *??religion history, telephone system* $\sim$ *??telephonic system*. This variation is in stark contrast to the productivity of genuine N2A transpositions.

[6]The Japanese nominalizer *-sa* may be a case in point (Sugioka, 2011).

morphology). On the other hand, participial adjectives such as *interesting* differ from ordinary property adjectives because they can be modified by words which modify the verb element of the participle: *a physically/mentally challenging task* = a task which challenges someone physically/mentally, *a partially broken chair* = a chair which has been broken but not completely, and soon. Even with ordinary gradable adjectives we find degree modifiers which do not modify a straightforward degree component but which entail a more complex kind of semantic modification: *a deceptively tall building, a suspiciously enthusiastic response*. Given this we can conclude that *very* in (10a) modifies the degree to which an entity interests another entity.

If this reasoning is correct then most of the participial lexicalized adjectives turn out to be transpositional in the sense that they inherit their basic meaning from (one) meaning of the base verb and do not add a semantic predicate. However, they are distinct (adjectival) lexemes so they do not inherit the full array of verbal properties from their base, and in particular cannot be modified in the manner of verbs (though they can take typical adjectival modifiers, with atypical meanings).[7]

6 Conclusions

I have argued that the model of lexical relatedness presented in Spencer (2013), Generalized Paradigm Function Morphology, is readily compatible with the LFG architecture. A central feature of that model is the deployment of a Lexemic Index (LI), which serves to individuate lexical entries/lexemes. That LI turns out to have essentially the same function as the naming component of LFG's PRED attribute.

One of the reasons why it has been so difficult for morphologists to distinguish inflection from derivation is the existence of transpositions such as deverbal participles, which appear to be forms of the base lexeme and fail to alter the semantic content of the base, just like canonical inflection, but which do change the morphological and syntactic category of that base, just like derivation. In the GPFM model relatedness is factorized so that it is not actually important to decide whether a type of relatedness is strictly inflectional, derivational, or some intermediate type such as a transposition. I also argue that the LI has to be an arbitrary label, such as a unique integer, rather than being identified with phonological or semantic representations. One of the reasons why the LI can't be identified with the lexeme's semantics is that there are transpositions such as the relational adjectives of Selkup, two of which do change the semantic representation, adding similitudinal and locational meanings,

[7] Spencer (forthcoming) proposes that the Principle of Representational Independence be restricted to just those cases in which the LI is changed. The proposal there is that where two words are related but have distinct LIs then they must also be distinct along some dimension of 'interpretation', i.e. either their SEM value has to be distinct or, failing that, their a-structure. This would then rule out a situation in which a language with very rich inflection has lexemes with thousands of different forms, each one of which could in principle be given its own LI. However, the revised statement of the PRI in Spencer (forthcoming) would also rule out pure synonymy, which may or may not be an advantage. I leave this to future research.

but which can be formed from nouns inflected for possessor agreement and which are thus themselves effectively inflected forms of the base noun lexeme, and not the product of derivational morphology.

On the other hand, the LI also allows us to distinguish the truly transpositional type of adjective from superficially similar cases such as English participial adjectives of the type *interesting*. These are like true (participial) transpositions in that they do not add a predicate to the SEM representation of the base verb, and yet they can be shown to behave like distinct, derived lexemes. They are thus instances of a 'transpositional lexeme', adding to the types identified in Spencer (2013). The LI, now PRED, attribute plays a crucial role in distinguishing this type from true transpositions, thus providing further motivation for retaining that attribute in the LFG architecture.

References

Alexiadou, Artemis. 2010a. Nominalizations: a probe into the architecture of grammar. Part I: the nominalization puzzle. *Language and Linguistics Compass* 4, 496–511.

Alexiadou, Artemis. 2010b. Nominalizations: a probe into the architecture of grammar. Part II: the aspectual properties of nominalizations, and the lexicon vs. syntax debate. *Language and Linguistics Compass* 4, 512–523.

Andrews, Avery. 2008. The role of PRED in LFG+glue. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG08 Conference*, pages 46–67, Stanford, CA: CSLI Publications.

Apresjan, Jurij Derenikovič. 1974. *Leksičeskaja semantika. Sinonimičeskie sredstva jazyka*. Moscow: Nauka.

Asudeh, Ash. 2012. *The Logic of Pronominal Resumption*. Oxford: Oxford University Press.

Bach, Emmon and Chao, Wynn. 2009. On Semantic Universals and Typology. In Morten H. Christiansen, Chris Collins and Shimon Edelman (eds.), *Language Universals*, pages 152–173, Oxford: Oxford University Press.

Bach, Emmon and Chao, Wynn. 2012. Semantic types across languages. In Claudia Maienborn, Klaus von Heusinger and Paul Portner (eds.), *Semantics. An International Handbook of Natural Language Meaning*, volume X, pages 2537–2558, Berlin: De Gruyter Mouton.

Bierwisch, Manfred. 1983. Semantische und konzeptuelle Repräsentationen lexikalischer Einheiten. In Rudolf Ružičká and Wolfgang Motsch (eds.), *Untersuchungen zur Semantik, Studia Grammatika 22*, pages 61–99, Berlin: Akademie Verlag.

Bonami, Olivier. 2015. Periphrasis as collocation. *Morphology* 25, 63–110.

Borer, Hagit. 2013. *Structuring Sense: Taking Form*. Oxford: Oxford University Press.

Corbett, Greville G. 2007. Canonical Typology, Suppletion, and Possible Words. *Language* 83(1), 8–42.

Croft, William. 2003. *Typology and Universals*. Cambridge: Cambridge University Press, second edition.

Cruse, Alan D. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.

Harley, Heidi. 2014. On the identity of roots. *Theoretical Linguistics* 40, 225–275.

Haug, Dan and Nikitina, Tanya. 2012. The many cases of non-finite subjects: the challenge of "dominant" participles. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG12 Conference*, pages 292–311, Stanford, CA: CSLI Publications.

Jackendoff, Ray S. 1990. *Semantic Structures*. Cambridge, MA: The MIT Press.

Koontz-Garboden, Andrew. 2007. *States, changes of state, and the Monotonicity Hypothesis*. Ph.D. thesis, Stanford University Department of Linguistics, Stanford, CA.

Koptjevskaja-Tamm, Maria. 1993. *Nominalizations*. London: Routledge.

Kuriłowicz, Jerzy. 1966. Dérivation lexical et dérivation syntaxique (contribution à la théorie des parties du discours). In Eric Hamp, Fred Householder and Robert Austerlitz (eds.), *Readings in Linguistics volume 2*, pages 42–50, Chicago: University of Chicago Press.

Lehmann, Christian. 1984. *Der Relativsatz: Typologie seiner Strukturen, Theorie seiner Funktionen, Kompendium seiner Grammatik*. Tübingen: Gunter Narr Verlag.

Levin, Beth and Rappaport Hovav, Malka. 2005. *Argument Realization*. Cambridge: Cambridge University Press.

Lieber, Rochelle. 2004. *Morphology and lexical semantics*. Cambridge: Cambridge University Press.

Malchukov, Andrej L. 2004. *Nominalization/verbalizations: constraining a typology of transcategorial operations*. Munich: LINCOM EUROPA.

Malchukov, Andrej L. 2006. Constraining nominalization: function-form competition. *Linguistics* 44, 973–1008.

Marantz, Alec. 1997. No escape from syntax: don't try morphological analysis in the privacy of your own lexicon. In A. Dimitriadis, L. Siegel, C. Surek-Clark and A. Williams (eds.), *University of Pennsylvania Working Papers in Linguistics*, volume 4, pages 201–225, Department of Linguistics, University of Pennsylvania.

Nikitina, Tatiana. 2008. *The Mixing of Syntactic Properties and Language Change*. Ph. D.thesis, Stanford University, Stanford, CA.

Pustejovsky, James. 1995. *The Generative Lexicon*. Cambridge, MA: The MIT Press.

Sag, Ivan A. 2012. Sign-Based Construction Grammar: an informal synopsis. In Hans C. Boas and Ivan A. Sag (eds.), *Sign-Based Construction Grammar*, pages 69–202, Stanford, CA: CSLI Publications.

Spencer, Andrew. 1999. Transpositions and argument structure. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1998*, pages 73–102, Dordrecht: Kluwer Academic Publishers.

Spencer, Andrew. 2005. Towards a typology of 'mixed categories'. In C. Orhan Orgun and Peter Sells (eds.), *Morphology and the Web of Grammar. Essays in Memory of Steven G. Lapointe.*, pages 95–138, Stanford, CA: CSLI Publications.

Spencer, Andrew. 2013. *Lexical Relatedness: A Paradigm-based Model*. Oxford: Oxford University Press.

Spencer, Andrew. This volume. Participial relatives in LFG.

Spencer, Andrew. Forthcoming. How to individuate lexemes. In Daniel Siddiqi and Heidi Harley (eds.), *Morphological Metatheory*, Amsterdam: John Benjamins Publishing Co.

Stump, Gregory T. 2001. *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

Stump, Gregory T. 2002. Morphological and syntactic paradigms: arguments for a theory of paradigm linkage. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 2001*, pages 147–180, Dordrecht: Kluwer Academic.

Stump, Gregory T. Forthcoming. *Inflectional Paradigms*. Cambridge: Cambridge University Press.

Sugioka, Yoko. 2011. Nominalization affixes and multi-modularity of word formation. In Etsuyo Yuasa, Tista Bagchi and Katharine Beals (eds.), *Pragmatics and Autolexical Grammar. In Honor of Jerry Sadock*, pages 143–162, John Benjamins Publishing Co.

Vendler, Zeno. 1968. *Adjectives and Nominalizations*. The Hague: Mouton de Gruyter.

Williams, Edwin. 1981. Argument structure and morphology. *The Linguistic Review* 12, 18–114.

Wunderlich, Dieter. 1997. Cause and the structure of verbs. *Linguistic Inquiry* 28, 27–68.

Zucchi, Alessandro. 1993. *The Language of Propositions and Events*. Dordrecht: Kluwer Academic Publishers.

PARTICIPIAL RELATIVES IN LFG

Andrew Spencer
University of Essex

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King (Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

I propose an analysis of the attributive participle (verb-to-adjective, V2A, transposition) functioning as a participial relative clause, PTCP-RC. I adopt Spencer's (2013) analysis of transpositions: the argument structure representation of a verb has an event 'EV' semantic function (SF) role, an adjective has a modifier 'A-MOD' SF role and a participle has a composite ⟨A-MOD⟨ EV⟩⟩ SF role. The composite SF role licences both verb and adjective morphosyntax through a projection function which maps a-structure representations to c-structure and f-structure. Specifically, the A-MOD SF role licences adjective-noun concord. I illustrate this with Lithuanian, whose participles have the full array of adjectival properties, including definiteness marking. I offer brief speculation on how the analysis might be extended to languages whose PTCP-RCs can be formed on non-subject GFs.

1 Introduction

In many languages verbs take participle forms, which have the outward appearance and function ('external morphosyntax') of attributive adjectives, but which retain crucial inherent properties of verbs ('internal morphosyntax').[*] For instance, a participle-headed phrase (PtcpP) may occupy the typical position of an adjective phrase and may agree with the head noun like an adjective, but it may take complements and be modified by adverbials just like the base verb and may express verb properties such as voice, aspect, tense or even mood. Such PtcpPs function exactly like relative clauses (RCs), for instance, the Relative Pronoun RCs (RelPron-RC) of English (Lehmann, 1984), and should be considered a subtype of relative clause, the Participial Relative Clause (PTCP-RC).

In LFG such constructions have recently been the subject of investigation for Latin (Haug and Nikitina, 2012, 2015) and Sanskrit (Lowe, 2015) (there is interesting, if brief, discussion too in Falk, 2006). Lowe presents a detailed analysis of early and Classical Sanskrit showing that much of the tense system of the finite verb is retained in the participle system. He therefore proposes to treat PtcpPs exactly like RelPron-RCs at f-structure. Haug and Nikitina, however, argue that the f-structure of the PtcpP is an XADJ whose SUBJ attribute is controlled by the head noun through cyclic re-entrancy. A question that is not discussed in any detail by either set of authors is the precise relation between the purely adjectival aspects of the morphosyntax of the participles and that of true adjectives.

I will argue for an analysis of such PtcpPs which more closely reflects their overlap with true adjectives, basing my discussion on the participial system of Lithuanian. This is almost identical in structure and function to its sister language, Sanskrit, but there are additional features bringing it closer to the adjectival system which call into question the approach taken by Lowe.

[*]I am grateful to Mary Dalrymple and the editors for helpful discussion which has eliminated various errors and improved the clarity of the exposition.

The main problem posed by PTCP-RCs is that they are mixed categories, in fact, transpositions (Kuriłowicz, 1966; Beard, 1995), akin to action nominalizations. Relatively little attention has been given to mixed categories in LFG, the principal recent sources being the treatment of agent nominals in Gĩkũyũ proposed by Bresnan (1997), Bresnan and Mugane (2006), and the discussion of action nominals in Nikitina (2008). There are technical and conceptual problems with Bresnan and Mugane's approach (Spencer, 2013). I therefore propose an analysis which combines LFG morphosyntax with the approach to lexical representation argued for in Spencer (2013).

2 Functions of participial phrases

Languages often use participial verb forms and whole participial clauses for purposes other than attributive modification, i.e. PTCP-RCs. In particular we often find that the participial clause functions more like a clause adjunct than the modifier of a noun. In some cases it is difficult to distinguish such uses from appositive or non-restrictive relative clauses (especially in Australian languages, for instance; see Nordlinger, 2014, 248f and references therein for recent discussion). Schematically, if we can treat a phrase such as [*the* [*girl* [*quietly reading a book in the corner*]]] as containing a (restrictive) RC, do we treat that clause as a non-restrictive RC or as a clausal adjunct in *The girl seemed wrapped in her own thoughts, quietly reading a book in the corner*? Languages seem to differ. Lowe (2015) discusses the matter in detail for Sanskrit and Haug and Nikitina (2012) discuss very interesting related constructions in Latin. I leave these constructions aside, however.

The canonical use of a participle form is as an attributive modifier, that is, PTCP-RC. There are broadly two types. In Indo-European languages the participle is effectively restricted to relativizing the SUBJ function (SUBJ-only PTCP-RCs). In other languages, however, we find constructions of the kind [*the boy* [*the girl giving* GAP *a book*]] or [*the girl* [GAP's *mother we meeting*], which we could call GF-PTCP-RCs (that is, relativizing on any grammatical function). This is the 'Option 1' type of Ackerman and Nikolaeva (2013). They violate the principle assumed in Haug and Nikitina (2012) and Spencer (2013) that the heads modified by participial clauses realize the subject of that clause.[1]

Many languages use participles in periphrastic tense-aspect-voice-polarity etc constructions with finite auxiliaries. We will see that Lithuanian is particularly rich in such constructions, but it also uses its participles as the sole exponent of an evidential mood series, without the mediation of any auxiliaries at all.

[1]Such PTCP-RCs are cases of head-marking. It is not the case that the participial morphology is really a clitic/edge inflection scoping over an entire phrase. For that kind of RC see, e.g., Zoque, (Faarlund, 2012, 158–169).

3 Participial relative clauses in Sanskrit

The PTCP-RCs of Vedic Sanskrit have been analysed within the LFG framework by Lowe (2015), one of the few in-depth studies in any framework of participial constructions. Lowe (2015: 79–83) effectively adopts the treatment of PTCP-RCs implicit in Bresnan and Mugane's (2006) discussion of Gĩkũyũ agent nominals. The participles are regarded as inflected forms of verbs bearing the attribute (VFORM)= PARTICIPLE. They differ from finite forms in that the participle is marked as non-finite and it specifies the gender of the SUBJ but not its person (while the finite form specifies person but not gender). In other respects the participles function very much like non-finite congeners of finite predicators.

In their function as attributive modifiers, participles fulfill the same function as Rel-PronRCs, and so Lowe proposes the same type of f-structure representation, in which the PTCP-RC is furnished with a REL-TOPIC attribute. The REL-TOPIC attribute bears the adjective-like agreement properties, mimicking adjective-like relative pronouns in languages such as Latin, Russian or Lithuanian.

The Sanskrit participles have exactly the same morphology and 'external syntax' (Haspelmath, 1996) of (a certain class of) adjectives, in particular, agreeing with the head noun exactly like an adjective does. When a participle get lexicalized as a true adjective it is the complementation and semantic properties that reveal this, not the agreement morphosyntax or morphological form. However, on Lowe's analysis, it is not clear how this convergence of morphosyntax with adjectives is to be represented, because the participle is presented as a type of verb which, for unclear reasons, takes adjective-type agreements. I will therefore present an alternative analysis under which the participle really is, morphosyntactically, an adjective (as well as being a verb).

4 Participial relatives in Lithuanian

The Lithuanian participle system is, if anything, more baroque than that of Sanskrit and it inherits a great many of the same Indo-European properties as the Sanskrit system (Ambrazas, 1997, henceforth 'LG'). Lithuanian finite verbs inflect for TENSE: {present, past simple, past frequentative, future}, ASPECT: {simple, perfect, progressive}, MOOD: {indicative, subjunctive, imperative, relative/oblique}, VOICE: {active, passive}. There is also a morphologically realized category of reflexive verb. Finite verbs agree with their subjects in person/number. Lithuanian non-finite forms are represented by an infinitive and a rich set of participles (*dalyviai*), gerunds (*padalyviai*), and the 'half-participles' (*pusdalyviai*, sometimes translated as 'semi-gerunds' or 'partly declinable' gerunds). The participles express the same range of four tenses as finite forms in active, passive and reflexive voice forms (together with modal variants, specifically a necessitive participle 'that which should be *VERB*-ed'). Like their Sanskrit congeners, the Lithuanian participles are morphologically adjectives and are fully incorporated into the adjectival

declensional class system, including the characteristic system of accentual classes. Lithuanian participials share with Sanskrit and other classical languages functions such as the gerund-like clausal adjunct and complements to certain sorts of predicate (e.g. perception verbs), which means that Lithuanian will function as a kind of proxy for Sanskrit.

In PTCP-RCs the complementation of the participles is that of the base verb, and participles are modified by (event-oriented) adverbials in the manner of verbs. The attributive usage of participles is described in LG:353–360. Typical examples are seen in (1–6), where the RC can be either prenominal (1–5) or postnominal (6) with no difference in interpretation as far as I am aware. (The past frequentative is not generally used attributively):

(1) Present active (PRS_ACT_PTCP):
darbus išman-ant-is ukininkas
jobs know-PRS_ACT_PTCP-M.SG.NOM farmer[M].SG.NOM
'a farmer knowledgeable about various jobs' LG:354

(2) Past active (PST_ACT_PTCP):
tolumoj pasirodęs žmogus
in.the.distance appear.PST_ACT_PTCP.M.SG.NOM man[M].SG.NOM
'the man who had appeared in the distance' LG:354

(3) Future active (FUT_ACT_PTCP, somewhat rare as a PTCP-RC):
Ateisiančios žiemos ilgumo
come.FUT_ACT_PTCP.F.SG.GEN winter[F].SG.GEN length
'the length of the coming winter' LG:355

The past passive participles (PST_PASS_PTCP) with perfective aspect are used as expected:

(4) pamiršta daina
forget.PST_PASS_PTCP.F.SG.NOM song[F].SG.NOM
'a forgotten song'

(5) Future passive participle (FUT_PASS_PTCP):
turėsimus pinigus
have.FUT_PASS_PTCP.F.ACC.PL money[F.PL].ACC
'money which he is going to have'

Present passive (PRS_PASS_PTCP; this form can express a wide variety of meanings):

(6) Durys visiems einamos
door[F.PL].NOM everybody.DAT walk.through.PRS_PASS_PTCP.F.PL.NOM
'a door everybody can walk through'

Participles in the nominative case can be used as open clausal adjuncts, controlled by the matrix SUBJ with which they agree in number/gender, and nominative case, in effect (these are the *pusdalyviai* 'half-participles').

(7) Mergaitės dainuodamos grėbe sieną
girls singing.F.PL.NOM mowed the-hay
'The girls mowed the hay (while) singing.'

Lithuanian also has a dative absolutive construction, a closed adjunct whose SUBJ has to appear in the dative.

(8) Jis išvažiavo lietui lyjant
he left rain.DAT raining
'He left while it was raining.'

The attributive and gerundial uses of the participles closely mirror the situation in Vedic Sanskrit. Lithuanian also uses participles together with forms of the verb BE to form compound constructions, much as in Classical Sanskrit and other later Indo-European languages. These include compound perfect and progressive tenses and the periphrastic passive construction. They are available for all tenses (including periphrastic ones), indicative/subjunctive mood, and active/passive voice. In (9) we see a 'nested' periphrasis: the passive is formed with the BE auxiliary and the passive participle, and that auxiliary itself appears in the present participle form with the (present) subjunctive of BE to form the past subjunctive.

(9) butume buvusi-os mušam-os
BE.SUBJ.1PL BE.PRS_PTCP-F.PL beat.PASS_PTCP-F.PL
'We[F] would have been beaten.'

However, Lithuanian participles are also used to realize a special relative or oblique mood, essentially a type of (hearsay) evidential (LG: 371).

(10) Seniau žmones nąmu nerakindavę
earlier people houses not.lock.PST_FREQ_PTCP.PL
'(I heard) People didn't lock their doors in olden times.'

(11) Jai nuo darbo rankas suką
her.DAT from work arms.[M] ache.PRS_PTCP-M.PL
'(She said) Her arms ache from work.'

(12) Kaip gražu budavę!
How beautiful be.PST_FREQ-PTCP
'How beautiful it used to be!'

(13) Kitąmet busią žąsų
next.year be.FUT_PTCP.M.PL geese
'Next year there'll be (probably) geese as well <i.e. just like this year>.'

The evidential meaning of the participle forms extends to the auxiliary BE in its use to form compound tenses (perfect, progressive).

In their use as the exponent of the evidential (relative) mood the participle forms look just like finite verb forms. However, they remain (predicative) adjectives in morphology and agreement properties. But this property highlights the conceptual problem with treating PTCP-RCs as essentially just another verb form. The participles in the relative mood form are the head of the main clause predication, not part of a subordinate clause. But if the PTCP-RCs of Lithuanian are to be treated as verb phrases then they will be headed by bare participles (lacking auxiliary verbs), precisely the construction that is normally interpreted as an evidential mood, yet they are interpreted as indicative mood forms, not relative mood. This is not because evidential modality is excluded from relative clauses: it is possible to put a relative clause formed using the RelPron-RC strategy in the relative/oblique mood, as seen in (14) (LG: 264):

(14) Jis mėgdavo svajoti apie veikalus, kuriu-os
he used.to.like to.daydream about books[F].GEN.PL REL_PRON-F.ACC.PL
jis parašiąs …
he write.FUT_PTCP …
'He liked to daydream about the books which he was going to write … [allegedly]'

5 Why participial relatives are adjective projections

PTCP-RCs are essentially non-finite verb phrases but in many languages they share a good deal of their morphosyntax with adjectives. To date, no account of PTCP-RCs that I am familiar with fully takes this fact into account.[2] In this section I outline the adjectival properties of a variety of PTCP-RCs in Lithuanian, where the parallels with adjectival morphosyntax are particularly close.

Lithuanian participles inflect exactly like adjectives (of a specific declension and accentual class), much like those of Sanskrit. However, Lithuanian adjectives in addition oppose an indefinite declension with a definite declension (LG: 142–147).

(15) a. GOOD: *geras* ∼ *gerasis*
M.SG.NOM M.SG.NOM.DEF

[2] A recent study of these constructions, Ackerman and Nikolaeva (2013), deals for the most part with languages (notably Tundra Nenets) which show little or no adjectival agreement, and in which word types described as 'participles' often behave more like (appositive) nominal phrases than adjectival phrases. Moreover, the focus of their study is the possessive relative clause, in which the head noun takes possessor agreement morphology cross-referencing the subject of the RC. For this reason they treat the PTCP-RCs as based on a generic 'mixed category' type, not a type of adjective. A detailed comparison between their HPSG-Construction Grammar approach and an LFG treatment would be very welcome.

b. BEAUTIFUL: *gražioms* ∼ *gražiesiems*
F.PL.DAT F.PL.DAT.DEF

The definite forms are obligatorily when there is a definite determiner in the NP, but on their own they can serve as the sole exponent of definiteness. Baltic languages lack a definite article. Definiteness is a property only of definite demonstratives, possessives and the like together with the definiteness inflection on an attributive adjective. Definiteness is not marked as such on RelPron-RCs (that is, the relative pronouns don't exhibit the definite ∼ indefinite contrast). Participles, however, inherit the definiteness category, as seen in (16).

(16) stovinči-o-ji moteris
stand-PRS_PRT-F.NOM.SG-DEF woman
'the standing woman'

Finally, one further argument from outside Indo-European suggests that we should encode adjectival properties of PTCP-RCs more explicitly than is possible if we assimilate the PTCP-RCs to RelPron-RCs by positing a REL-PRON attribute at f-structure. Cross-linguistically the PTCP-RCs are distinct from other morphologically expressed RCs. This can be seen from Swahili (Bantu [G42]). In (17) we see typical subject RCs (Ashton, 1944, 111).

(17) Mtu a-na/li/taka-**ye**-soma
person SM-PRS/PST/FUT-YE-read
'a person who is reading/read/will read'

a. Mtu a-soma-**ye**
person SM-read-YE
'a man who reads'

b. Mtu a-ki-soma-**ye** kitabu hiki
person SM-OM-read-YE book this
'a person who reads this book'

The RC is formed using the 'O-of-Reference' (Ashton, 1944), a multifunctional verb affix. The verb form is still finite, however, and shows no adjectival properties whatever, in particular, it shows the subject agreement morphology of a verb, not that of an adjective. These verb forms are therefore not participles. The kinds of f-structure representations appealed to in Lowe's account might well be suited to the Swahili relatives, but that just highlights the fact that they are inappropriate for the adjectival PTCP-RCs.

6 GPFM/LFG analysis of PTCP-RCs

6.1 Lexical categories in GPFM

My analysis will crucially depend on the treatment of transpositional morphosyntax first proposed in Spencer (1999) and elaborated within the Generalized Paradigm

Function Morphology (GPFM) model of Spencer (2013). First, I assume, following many authors, that lexical representations include a representation of argument structure (ARG-STR), distinct from a semantic role representation and from grammatical functions such as subject and object. I also assume that ARG-STR representations include what I will call 'semantic function (SF) roles': 'REF' (nouns), 'EV' (verbs). For adjectives I argue that the SF role is one which mediates attributive modification, A-MOD (A* in Spencer, 1999, 2013; see also Spencer, this volume[3]) Thus, the ARG-STR representation for TALL, PROUD(-OF) will be *tall*⟨A-MOD⟨x⟩⟩, *proud*⟨A-MOD⟨x,y⟩⟩. The usage of A-MOD is shown schematically in (18).

(18) **Swedish**⟨A-MOD$_x$⟨x⟩⟩ **man** ⟨REF$_x$⟩ ≈ λx(**Swedish**(x) ∧ **man**(x))

The notation REF$_x$ indicates that the 'REF' SF role is coindexed with the A-MOD role of the adjective, with the 'x' subscript on A-MOD, REF indicating that the adjective modifies the entity of which 'man' is predicated.

A transposition is a switch in the morphosyntactic category of a lexeme which is not associated with a change in the semantic representation of that lexeme. This is represented by overlaying an additional SF role in the argument structure, a morphological, lexical operation, which nonetheless has repercussions for syntax. In the case of the participle, we derive a representation category A-MOD from one of category EV, as shown in (19).

(19) ⟨EV⟨… x…⟩⟩ ⇒ ⟨A-MOD$_x$⟨EV⟨…x…⟩⟩⟩

In the case of (19) the referent of the noun (REF$_x$ element) that gets modified by the A-MOD category will also be identified with some thematic argument ('x') from the base verb's thematic array.

For languages which are very conservative with respect to the Keenan-Comrie (Keenan and Comrie, 1977) hierarchy for PTCP-RCs (e.g. Indo-European) the A-MOD SF is constrained to co-index only the SUBJECT (cf. Haug and Nikitina, 2012, on Latin participles). However, for languages which have PTCP-RCs on OBJ roles, or roles further down the Keenan-Comrie hierarchy the A-MOD SF is permitted to co-index any participant role on a language-specific basis. I return to this briefly in Section 6.2

A participle, then, is 'of the category' ⟨A-MOD⟨EV⟨…⟩⟩. On a language-specific basis the 'EV' and ⟨…⟩ components are able to license c-structures and f-structure attributes corresponding to the purely verbal categories. For Lithuanian, Sanskrit, and other languages this even extends to details of TAM semantics (Lowe 2015).

[3]I slightly modify the notation in this paper, omitting the '*' device, to avoid potential confusion with the Kleene-star.

6.2 LFG analysis

In lexical representations I assume an ARG-STR attribute which consists of an SF role which has at least two sets of values. The first is the lexical item's Lexemic Index (LI), essentially the 'name' component of the PRED attribute, while the second is a participant role. For prepositions and verbs the participant roles correspond to grammatical functions; for adjectives the principal participant role corresponds to the noun modified by the adjective; for nouns the participant role corresponds to the denotation of the noun itself.

(20) noun:

$$\begin{bmatrix} \text{REF} \begin{bmatrix} \text{LI} & \text{man} \\ \text{ARG1} & [\ ] \end{bmatrix} \end{bmatrix}$$

cf. λx.**man**(x), where x ≋ ARG1 [ ]

(21) verb:

$$\begin{bmatrix} \text{EV} \begin{bmatrix} \text{LI} & \text{'read'} \\ \text{ARG1} & [\ ] \\ \text{ARG2} & [\ ] \end{bmatrix} \end{bmatrix}$$

cf. λx,y.**read**(x,y)

(22) attribute (transitive) adjective:

$$\begin{bmatrix} \text{A-MOD} \begin{bmatrix} \text{LI} & \text{'proud(-of)'} \\ \text{ARG1} & [\ ]_i \\ \text{ARG2} & [\ ] \end{bmatrix} \end{bmatrix}$$

cf. λx,y.**proud**(x,y)

The 'i' subscript in (22) indicates that the ARG1 is shared with the ARG1 of the head noun (see below):

(23)

$$\begin{bmatrix} \text{REF} \begin{bmatrix} \text{LI} & \text{'man'} & \\ \text{ARG1} & [\ ]_i & \\ \text{A-MOD} & \begin{bmatrix} \text{LI} & \text{'tall'} \\ \text{ARG1} & [\ ]_i \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

cf. λx.[**tall**(x) ∧ **man**(x)]

A participle is defined by embedding a EV structure under the A-MOD structure:

(24)
$$\begin{bmatrix} & \begin{bmatrix} LI_j & \\ ARG1 & [\,]_i \end{bmatrix} \\ A\text{-}MOD & \begin{bmatrix} & \begin{bmatrix} LI & \text{'read'}_j \\ ARG1 & [\,]_i \\ ARG2 & [\,] \end{bmatrix} \\ EV & \end{bmatrix} \end{bmatrix}$$

cf. λx,y.**reading**(x,y)

Note that the ARG1 (subject argument) of the verb is shared with the adjectival ARG1. In this way the head noun will (ultimately) be identified as the SUBJ of the participle. Note too that the LI value of the A-MOD is shared with that of the verb (EV), because the participle is a form of the verb lexeme and not a distinct lexeme in its own right.

The participle can serve as the modifier of a noun in just the same way that an ordinary adjective can. Hence, at ARG-STR we have (25):

(25)
$$\begin{bmatrix} REF & \begin{bmatrix} LI & \text{'girl'} \\ ARG1 & [\,]_i \end{bmatrix} & \\ A\text{-}MOD & \begin{bmatrix} LI & [\,]_j & \\ ARG1 & []_i & \\ EV & \begin{bmatrix} LI & \text{'read'}_j \\ ARG1 & [\,]_i \\ ARG2 & [\,] \end{bmatrix} \end{bmatrix} & \end{bmatrix}$$

I now consider how an eventive predicate embedded under an adjectival A-MOD argument structure is represented at f-structure and at c-structure, beginning with the mapping to f-structure.

To begin with, following Spencer (this volume), I assume that traditional LFG PRED values are fractionated as follows: for all parts of speech, the PRED FN or 'name' part of the PRED value (for instance, the component 'see' in [PRED 'see⟨ SUBJ, OBJ⟩']) is equivalent to GPFM's Lexemic Index (LI). For any part of speech that has a non-trivial argument structure which is expressed as an array of grammatical functions, we separate out that GF array as a separate attribute. Thus, [PRED 'see⟨ SUBJ, OBJ⟩'] becomes (26).

(26)
$$\begin{bmatrix} LI & \text{'see'} \\ GF\text{-}ARRAY & \langle SUBJ, OBJ \rangle \end{bmatrix}$$

Clearly, this notational convention has no other effect on our analysis.

For nouns we therefore generally assume an f-description limited to the LI so that the a-structure representation [REF [LI 'girl']] will correspond to the f-structure [LI 'girl']. Some nouns, such as kin terms and meronyms, have a richer argument

structure, as seen in (27), as is also the case with nouns that take clausal complements (*fact, idea, opinion, ...*). Deverbal nominalizations will often inherit (much of) the base verb's argument structure, as will certain deadjectival nominals (*pride in NP, difficulty with, certainty that*).[4]

(27) $\begin{bmatrix} & \begin{bmatrix} \text{LI} & \text{'daughter/hand'} \\ \text{ARG1} & [\ ] \\ \text{ARG2} & [\ ] \end{bmatrix} \\ \text{REF} & \end{bmatrix}$

A verb's f-description is principally defined by its GF-ARRAY, which can generally be defined straightforwardly from the ARG-STR representation.

For an adjective we need to map the ARG-STR representation to the f-structure of an ADJUNCT. This will depend on exactly how we view the f-structure of an attributive adjective. I will follow Haug and Nikitina (2012) in assuming that adjectives define open XADJ f-structures bearing a SUBJ attribute re-entrant with the f-structure of the whole phrase (see (30) below).

Assuming a language with SUBJ-only PTCP-RCs, a participle such as 'reading' has the ARG-STR representation:

(28) $\begin{bmatrix} & \begin{bmatrix} \text{LI} & [\ ]_i & \\ \text{ARG1} & [\ ]_j & \\ & \begin{bmatrix} \text{LI} & \text{'read}_i\text{'} \\ \text{ARG1} & [\ ]_j \\ \text{ARG2} & [\ ] \end{bmatrix} \\ \text{EV} & \end{bmatrix} \\ \text{A-MOD} & \end{bmatrix}$

The participle's '[ARG1 [ ]$_j$]' attribute is shared with the embedded EV ARG1 attribute. The participle's '[ARG1 [ ]$_j$]' attribute is also shared with the LI of the noun antecedent which it modifies, as indicated by the subscripting.

The 'EV' component of the participle's ARG-STR representation licenses a GF-ARRAY and hence, at c-structure licenses a VP (with OBJ, but no SUBJ) or a clause-like structure, ⟨SUBJ, OBJ⟩. The f-descriptions (and a-descriptions) define the correspondence between ARG2 and the OBJ attribute in the PTCP-RC's f-structure.

I have rejected an f-structure analysis of attributive participles as equivalent to RelPron RCs. One of the reasons is because the agreement morphosyntax of the participle is identical to that of an attributive adjective and this is not captured if the PTCP-RC is treated as bearing a covert relative pronoun. For simplicity, I will adopt the proposal of Haug and Nikitina (2012) in their discussion of the morphosyntax of adnominal participles in Latin. These, like the participles of Sanskrit and Lithuanian, double as adjuncts and as complement clauses to verbs of perception and can also be used predicatively and in periphrastic constructions. Haug

[4]Expletives require their own LI even though they have no semantic representation (see also Spencer, this volume).

and Nikitina adopt an attribute AGR which bundles together the agreement features for adjective-noun concord. They assume that the attributive participle is an open adjunct to the noun, whose SUBJ attribute is identified with that of the head noun itself, giving a cyclic f-structure. The participle itself is given its own AGR attribute. This is required in order to account for the unusual mismatched agreement configuration in the 'dominant' construction, illustrated in (29) (from Tacitus' *Annals*, adapted from Haug and Nikitina, 2012: 295).

(29) occisus dictator
kill.PST_PASS_PTCP.NOM.M.SG dictator[M].NOM.SG
Caesar … pulcherrimum facinus
Caesar[M].NOM.SG most.glorious.N.NOM.SG deed[N].NOM.SG
uideretur
seemed
'the slaying of Dictator Caesar seemed … the most glorious deed.'

The point of (29) is that the past passive participle *occisus* is not to be interpreted as an attributive modifier of the syntactic head *dictator Caesar* ('the slain Dictator Caesar'), but as an eventive deverbal nominalization with *dictator Caesar* as the object of the implied base verb *occido* 'kill'.

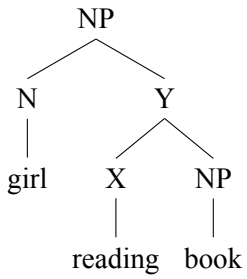
As an instance of standard attributive participial use Haug and Nikitina cite the Latin example *rosa florens*, rose bloom.PRS_PTCP, 'a blooming rose'. On their account, this would have the f-structure shown in (30).

(30)
$$\begin{bmatrix} \text{PRED} & \text{'rose'} & \\ \text{AGR} & \begin{bmatrix} \text{CASE} & \text{NOM} \\ \text{GENDER} & \text{FEM} \\ \text{NUMBER} & \text{SG} \end{bmatrix} & \\ \text{XADJ} & \begin{bmatrix} \text{PRED} & \text{'bloom}\langle\text{SUBJ}\rangle\text{'} \\ \text{AGR} & [\ ] \\ \text{SUBJ} & \begin{bmatrix} \dots \\ \text{AGR} & [\ ]_{i} \end{bmatrix} \end{bmatrix}_{i} & \end{bmatrix}$$

We now simply have to state that it is a property of any category whose 'external' SF role is A-MOD that it has the f-structure AGR attribute. In this way we capture the generalization that is missed when participles are treated as having a covert relative pronoun.

I now turn to the mapping to c-structure. Consider our schematic example of *(a) girl reading (a) book*. The simplest c-structure would be that in (31).

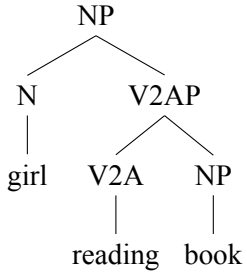
(31)



What is the identity of nodes X, Y? The obvious assumption is that Y is the projection of the head, X, hence, XP. What is the identity of X?

The most straightforward assumption is that X is the c-structure correspondent of a transposition, with ARG-STR representation [A-MOD [EV…]]. The external A-MOD SF role ensures that *reading* has the syntactic privileges of occurrence of an adjective. However, the embedded/internal EV SF role allows X to govern a direct object, too. Following Spencer (this volume) I will label this X node 'V2A' (for 'verb-to-adjective transposition') and so XP will be V2AP. This is an instance of a single-category approach to mixed categories (somewhat similar to the analysis of mixed categories proposed by Lapointe, 1993).

(32)



The PTCP-RCs which relativize on a non-SUBJ GF will require a different set of mappings from a-structure to f-structure, so as to permit the base verb's direct object to be linked to the A-MOD[ARG1] role, and not just the subject, schematically: *the book* [*the girl quietly reading* __ *in the corner*].

In a language with GF-PTCP-RCs we need to assume as many ways of specifying the ARG1 argument of the participle as there are possible relative clause gap types. In the case of a language which permits, say, just relativization on subjects and objects, either the participle's [ARG1 [ ]$_i$] is shared with the EV attribute's ARG1 position, as in SUBJ-only PTCP-RCs or it is shared with ARG2. In the latter case, the OBJ role of the base verb corresponds to the RC gap: if the language has a VP, the c-structure will include a VP(-like) phrase but it will lack a NP OBJ correspondent. On the other hand, it may have a overt NP SUBJ correspondent. This means that the (default) mapping which defines the ARG1 of the EV component will apply, because it is not overridden.

The kinds of non-SUBJ RCs that need to be accounted for are illustrated from

Turkish in (33–36) (adapted from Göksel and Kerslake, 2005, 438).[5] Turkish lacks agreement, so the f-structures are somewhat simpler than in the case of Sanskrit, Lithuanian, or Latin.

(33) oyuncak-lar-ın-ı kır-an küçük kız
toy-PL-3SG.POSS-ACC break-PTCP little girl
'the little girl who breaks/has broken her toys' [SUBJ gap]

(34) her gün okul-da gör-duğ-üm kız
every day school-LOC see-PTCP-1SG.POSS girl
'the girl whom I see at school every day' [OBJ gap]

(35) anne-si-yle tanış-acağ-ım kız
mother-3SG.POSS-COM meet-PTCP-1SG.POSS girl
'the girl whose mother I'm going to meet' [POSS gap]

(36) baş-ın-da şapka ol-an kız
head-3SG.POSS-LOC hat be-PTCP girl
'the girl who has a hat on her head' [POSS gap]

The kind of f-structure I envisage for such cases is illustrated in (37) for (35).

(37)
$$\begin{bmatrix} \text{PRED} & \text{'girl'} & \\ \text{XADJ} & \begin{bmatrix} \text{PRED} & \text{'meet}\langle\text{SUBJ,OBJ}\rangle\text{'} \\ \text{SUBJ} & \text{'I'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'mother'} \\ \text{POSS} & [\ ] \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The grammar of relative clauses in languages such as Turkish therefore has to furnish appropriate equations to mandate control of the object's possessor function by the head, as well as other non-subject grammatical functions.

Languages differ in the way they permit an overt SUBJ to be expressed in a PTCP-RC relativizing on a non-subject. In Tundra Nenets we see the SUBJ appearing in genitive case and also possessor marking on the head noun, cross-referencing that SUBJ (Ackerman and Nikolaeva's 'Option 2'). In other languages, we see just genitive case marking of the SUBJ, without possessor agreement ('Option 1'). In absolute constructions (that is, not attributive PTCP-RCs proper) we see the SUBJ expressed in various oblique cases (for instance, in Sanskrit: locative/genitive, Ancient Greek: genitive, Latin: ablative, Old Church Slavonic/Lithuanian: dative). However, in other languages the SUBJ of a PTCP-RC can be expressed just as it is in a finite clause. This is seen in the Kiranti language, Limbu, where the SUBJ

[5] The morphosyntax of non-subject PTCP-RCs requires a separate study, so I will not pursue these questions here. A very brief sketch of an LFG analysis of Turkish participial relative clauses is presented in Çetinoğlu and Oflazer (2006, 2009), but it does not address the issues of concern here.

is marked absolutive (38) or ergative (39) depending on the relative clause's transitivity.[6]

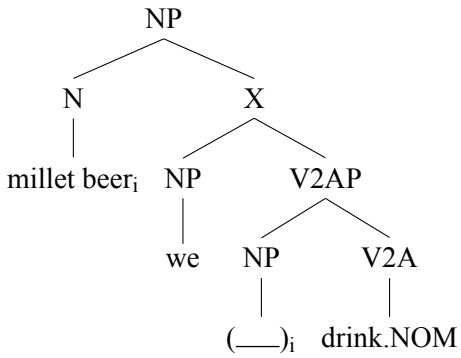
(38) anchige thuŋɛtchugebɛn ti·
we drank.PTCP.3.PATIENT.NOM.ABS millet.beer
kudzaphɛʔr-ɛ
taste.bad-PT
'The millet beer we drank tasted bad.'

(39) aŋga a-mma-re thapt-aŋ-ba ku-saʔ-ʔɛ
I my-mother-ERG bear-PTCP.1SG.NOM her-son-am
'I am [her] son that my mother bore./I am the son my mother bore.'

This raises the question of what is the identity of the node 'X?' in (40, linear order of V2A, NP in V2AP irrelevant),

An approximation to the c-structure of (38) (assuming that the subject position is external to the 'VP' node) is given in in (40) (linear order of V2A, NP in V2AP irrelevant). Again, we need to determine the identity of the node X in (40).

(40)



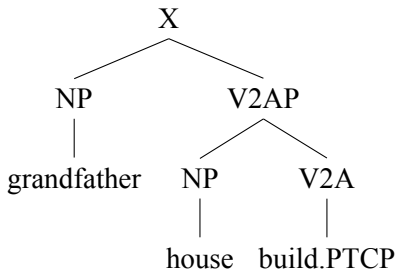
A clue may lie in the participial strategy often used for internally headed RCs, as seen in (41) from the Tungusic language, Udihe (Nikolaeva and Tolskaya, 2001), whose constituent structure is shown in (42), schematically represented in (43).

(41) Bi od'o-i zugdi-we wo:n-di-nii bi bagdi-mi
1SG grandfather-1SG house-ACC make.PRS_PTCP-DAT-3SG 1SG live-1SG
'I live in the house which my grandfather built.'

(42) [[Bi od'o-i] zugdi-we wo:n-di-nii] bi bagdi-mi

[6]The examples are (32) and (39) respectively from van Driem (1987, 196, 197) respective. I have adapted the glossing for the sake of ease of exposition.

(43)



X is in one sense clausal, but it is not a finite clause and it has the function of an attributive modifier. The simplest solution might therefore be to treat X as a projection of V2AP (in the same sense that S can be thought of as a projection of VP). However, the final decision will depend on the specific facts of each language and it will also depend on exactly how we analyse the full array of non-subject participial relative clauses, such as those found in the Turkic languages and those surveyed in Ackerman and Nikolaeva (2013), especially those that realize verbal tense-aspect-mood-voice categories.

7 Excursus: Participial modifiers as extended V heads

In this excursus I briefly consider the possibility of deploying an 'extended head' analysis to PTCP-RCs. Bresnan and Mugane (2006), following Bresnan (1997), propose an analysis of the Gĩkũyũ agent nominal mixed category, deploying the notion of extended head.[7] The agent nominal is derived morphologically from a verb, much like English *-er* nominals, but although the nominal behaves as though it heads a noun phrase with respect to determiners and other nominal modifiers, it also behaves as though it heads a verb phrase with respect to complements and event-oriented modifiers (adverbials). The key example is (44).

(44) ũyũ mũ-thĩĩnj-i mbũri ũũru
1.DEM 1-slaughter-NOM 10.goat badly
'this bad goat slaughterer'

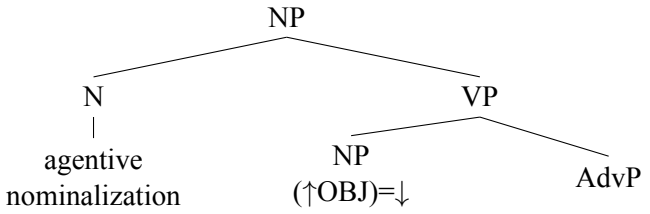
In (44), *mbũri* 'goat' is a direct object, not a complement to a (canonical) noun, and *ũũru* 'badly' is a verb modifier, not an adjective, yet the head, *mũthĩĩnji*, is a noun of Class 1, determined by the (agreeing) demonstrative adjective *ũyũ*. We thus have a case of category mixing comparable to an English example such as *Tom's singing the song loudly*.

The Bresnan and Mugane treatment is discussed in Spencer (2013, 254–256), where it is compared with the kind of approach sketched in this paper. Here I will explore the possibility of extending extended heads to PTCP-RCs.

The c-structure which Bresnan and Mugane assume for the agent nominal is that shown in (45) (cf. Bresnan and Mugane's example (74)):

[7] One can think of the analysis proposed in this paper as cashing out the suggestions made in Bresnan (1997, 14–15), on the basis of an early version of Spencer (1999).

(45) Extended head analysis of agentive nominalization



The agentive nominalization *mũthĩĩnji* is an autonomous word form which therefore occupies a single syntactic terminal. The VP node dominates the various complements and adjuncts that we would normally associate with a fully-fledged verb, but that VP node does not actually dominate a V, rather, the deverbal nominalization under the N node serves as the extended head of the VP node. VP appears under NP and not the other way around by virtue of extended head theory.

To get the analysis to work Bresnan and Mugane need to keep track of the 'derivational history' of the nominalization. They do this by invoking a subscript notation, shown in (46) (Bresnan and Mugane's example (72), p. 227).

(46) mũthĩĩnji: 'agent-of $\langle x$, slaughter$\langle x,y\rangle_v\rangle_n$'

They then propose the lexical entry in (47) (their example (73), p. 228) for the agentive nominal.[8]

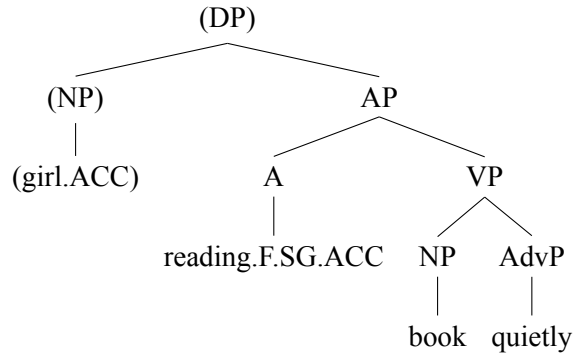
(47) mũthĩĩnji: N: (↑ PRED = 'slaughterer$\langle\langle$(↑ OBJ)$\rangle_v\rangle_n$'
v: VP ∈ CAT(PRED↑)
n: NP ∈ CAT(PRED↑)

If we were to adopt such an approach for the PTCP-RCs then the verb-oriented component of the RC would be represented in c-structure as a VP, while the adjectival external syntax of PTCP-RC as a whole would be reflected in an AP c-structure, as shown schematically in (48). The 'v', 'a' subscripts in (49) correspond to the notational innovations introduced by Bresnan and Mugane to allow the grammar to keep track of the fact that the mixed category (here, the participle) has to be associated both with a VP and an AP c-structure correspondent.[9]

[8]Mary Dalrymple points out to me that the f-descriptions provided by Bresnan and Mugane for the '*v, n*' subscripts in (47) should read *v*: VP ∈ CAT(↑), *n*: NP ∈ CAT(↑) respectively.

[9]This analysis seems to imply that 'reading' is a different lexeme from 'read' because it has a different PRED value. However, this can't be right (and probably is not right for Gĩkũyũ either). No such consequence follows in the GPFM analysis proposed above.

(48) ‘Extended head’ approach extended to participial attributes:



(49) F-structure:
a. read: V: (↑PRED) = ‘read⟨(↑SUBJ)(↑OBJ)⟩$_v$’
b. reading: A: (↑PRED) = ‘reading⟨⟨(↑SUBJ) (OBJ)⟩$_v$⟩$_a$’
v: VP ∈ CAT(PRED↑)
a: AP ∈ CAT(PRED↑)

This type of analysis seems to be feasible for the Indo-European SUBJ-only-PTCP-RC type. However, it is unclear how well it would work for PTCP-RCs more generally, specifically, those in which the relative clause targets a non-subject. The problem is that the c-structure category of the PTCP-RC has to be that of a (real) clause and not just a VP. Yet at the same time, that clause would have to be headed by an adjective. It is for this sort of reason that I argue that we should treat c-structure labels as projections from ARG-STR, specifically SF role, labels.

8 Conclusions

I have argued for an approach to participial relative clauses which treats them on a par with adjectival phrases. This requires us to have an explicit treatment of verb-to-adjective transpositions. I adopt the proposals of Generalized Paradigm Function Morphology (Spencer, 2013, extended to LFG in Spencer, 2015). I assume that argument structure representations for lexemes include a semantic function role attribute, REF (noun), EV, (verb), A-MOD (for attributive adjective). A transposition has a complex semantic function role, so that a participle is (schematically) ⟨A$_x$⟨E…⟩⟩. In Indo-European languages such as Russian, Latin, Sankskrit, Lithuanian, the participle modifies the head noun in the manner of an adjective, including agreement.

I propose that we define mappings from the articulated a-structure lexical representations of GPFM to f-structures and c-structures. This brings several benefits. First, the f-structure of a participial relative is essentially a combination of a clause and a true adjective: there is no need to postulate a covert relative pronoun. Second, we can define the c-structure node labels directly from the SF roles of a-structure.

This means that we obviate the notorious problems of what c-structure category a mixed category such as a participle 'really' is. I conclude with a brief excursus, suggesting that an extended head analysis might be appropriate for the subject-only type of participial relative. On that analysis the relative clause would not be mixed at c-structure but would be an adjective phrase whose head can select a VP complement. However, it is unlikely that this analysis could be extended to participial relatives on non-subjects.

References

Ackerman, Farrell and Nikolaeva, Irina A. 2013. *Descriptive Typology and Linguistic Theory. A Study in the Morphosyntax of Relative Clauses*. Stanford: CSLI Publications.

Ambrazas, Vytautas (ed.). 1997. *Lithuanian Grammar*. Vilnius: Baltos lankos publishers.

Ashton, E. 1944. *Swahili Grammar*. London: Longman.

Beard, Robert. 1995. *Lexeme-Morpheme Base Morphology*. Stony Brook, NY: SUNY Press.

Bresnan, Joan. 1997. Mixed Categories as Head Sharing Constructions. In Miriam Butt and Tracy Holloway King (eds.), *The Proceedings of the LFG '97 Conference*, University of California, San Diego.

Bresnan, Joan and Mugane, John. 2006. Agentive nominalizations in Gĩkũyũ and the theory of mixed categories. In Miriam Butt, Mary Dalrymple and Tracy Holloway King (eds.), *Intelligent Linguistic Architectures: Variations on Themes by Ronald M. Kaplan*, pages 201–234, Stanford, CA: CSLI Publications.

Çetinoğlu, Özlem and Oflazer, Kemal. 2006. Morphology-syntax interface for Turkish LFG. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL*, pages 153–160.

Çetinoğlu, Özlem and Oflazer, Kemal. 2009. Integrating derivational morphology into syntax. In Nicolas Nicolov, Galia Angelova and Ruslan Mitkov (eds.), *Recent Advances in Natural Language Processing V*, pages 155–170, Amsterdam: John Benjamins Publishing Co.

van Driem, George. 1987. *A Grammar of Limbu*. Berlin: Mouton de Gruyter.

Faarlund, Jan Terje. 2012. *A Grammar of Chiapas Zoque*. Oxford: Oxford University Press.

Falk, Yehuda N. 2006. On the representation of case and agreement. In Miriam Butt and Tracy Holloway King (eds.), *The Proceedings of the LFG06 Conference*, pages 184–201, Stanford, CA: CSLI Publications.

Göksel, Aslı and Kerslake, Celia. 2005. *Turkish. A Comprehensive Grammar*. London: Routledge.

Haspelmath, Martin. 1996. Word-class changing inflection and morphological theory. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1995*, pages 43–66, Kluwer Academic Publishers.

Haug, Dag and Nikitina, Tanya. 2012. The many cases of non-finite subjects: the challenge of "dominant" participles. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG12 Conference*, pages 292–311, Stanford, CA: CSLI Publications.

Haug, Dag and Nikitina, Tatiana. 2015. Feature sharing in agreement. *Natural Language and Linguistic Theory* 33.

Keenan, Edward and Comrie, Bernard. 1977. Noun Phrase Accessibility and Universal Grammar. *Linguistic Inquiry* 8, 63–99.

Kuriłowicz, Jerzy. 1966. Dérivation lexical et dérivation syntaxique (contribution à la théorie des parties du discours). In Eric Hamp, Fred Householder and Robert Austerlitz (eds.), *Readings in Linguistics volume 2*, pages 42–50, Chicago: University of Chicago Press.

Lapointe, Steven G. 1993. Dual lexical categories and the syntax of mixed category phrases. In Andreas Kathol and Michael Bernstein (eds.), *Proceedings of the 11th Meeting of the Eastern States Conference on Linguistics*, pages 199–210.

Lehmann, Christian. 1984. *Der Relativsatz: Typologie seiner Strukturen, Theorie seiner Funktionen, Kompendium seiner Grammatik*. Tübingen: Gunter Narr Verlag.

Lowe, John. 2015. *Participles in Rigvedic Sanskrit. The Syntax and Semantics of Adjectival Verb Forms*. Oxford: Oxford University Press.

Nikitina, Tatiana. 2008. *The Mixing of Syntactic Properties and Language Change*. Ph. D.thesis, Stanford University, Stanford, CA.

Nikolaeva, Irina A. and Tolskaya, Maria S. 2001. *A Grammar of Udihe*. Berlin: Mouton de Gruyter.

Nordlinger, Rachel. 2014. Constituency and grammatical relations in Australian languages. In Harold Koch and Rachel Nordlinger (eds.), *The Languages and Linguistics of Australia: a Comprehensive Guide*, pages 215–262, Berlin: Walter de Gruyter GmbH & Co.

Spencer, Andrew. 2013. *Lexical Relatedness: A Paradigm-based Model*. Oxford: Oxford University Press.

MODELING THE COMMON GROUND FOR DISCOURSE PARTICLES

Mark-Matthias Zymla, Maike Müller and Miriam Butt
University of Konstanz

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King (Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

German modal or discourse particles make a contribution to the illocutionary force of an utterance. In this paper we integrate a model of belief sets held as part of the Common Ground into the Abstract Knowledge Representation (AKR) component of the German ParGram grammar. We follow recent accounts that divide the Common Ground into individual belief state sets for each of the discourse participants. We extend the existing German ParGram grammar to be able to parse (short) dialogs and to provide "pragmatic parsing" that integrates information about the belief state of discourse participants. Within this architecture, we provide an analysis of discourse particles that is theoretically well-motivated and can simultaneously be employed in an applied computational context.

1 Introduction

This paper presents an augmentation of the existing Abstract Knowledge Representation (AKR) integrated into the ParGram architecture (Bobrow et al. 2007). The augmentation provides a modeling of the Common Ground (CG) (Gunlogson 2002, Stalnaker 2002). Through this augmentation, we are able to model belief states and allow for the computational drawing of inferences based on belief states.

We illustrate our computational modeling of the CG via an analysis of German discourse particles, which are known to contribute to the illocutionary force of an utterance. A subset of the about 20 discourse particles used in German has been analyzed as signaling information that is pertinent to the CG. Consider (1), for example, where the discourse particle *ja* serves to establish or reconfirm the proposition p as being part of the CG (Lindner, 1991, Kratzer, 1999, Zimmermann, 2011). In (1) the speaker wants to establish that the dog in question is a watch dog (rather than, for example, a hunting dog).

(1) Das ist **ja** ein Wachhund.
That is **yes** a watch dog
'This is a watch dog (as I want to establish).'

Within LFG, the only existing treatment of German discourse particles that we are aware of is their implementation as part of the German ParGram grammar (Dipper, 2003), where they are analyzed syntactically as adverbs and functionally as adjuncts. Our interest is concentrated on the interface between syntax and semantics/pragmatics and we thus extend the existing ParGram implementation. This extension essentially provides a parser capturing formal pragmatic intuitions about discourse analysis: a pragmatic parser.

[†]This work is part of two ongoing research projects, *Tense and Aspect in Multilingual Semantic Construction* funded by the Nuance Foundation and *VisArgue* financed by the BMBF. For more information about the projects see http://ling.uni-konstanz.de/pages/home/partma/ and http://visargue.uni-konstanz.de. We thank the LFG15 audience for valuable feedback and Maribel Romero for continuous feedback and discussions. Tracy King was instrumental in helping us with the implementation.

The pragmatic parser uses syntactic and semantic information to produce a discourse model. The first step in devising this model is to organize the beliefs of the discourse participants (who believes what). We adopt the set-theoretic assumptions of Gunlogson's (2002) work. Throughout the discourse parsing, these beliefs are manipulated and changed with each new contribution to the discourse. Thus, a crucial task is to determine the intention of a certain contribution or utterance. Our current approach to this is to assume that utterances represent a small set of discourse moves such as rejection or acceptance. Our model allows for a check for logical consistency among discourse moves, an essential component for the analysis of discourse moves (Walker, 1996).

The German ParGram grammar already contains a basic semantic component (Zarrieß 2009a,b) that is realized within the AKR system. AKR has been used primarily for computational purposes such as Question-Answering (Q&A) systems (Condoravdi et al. 2003, Bobrow et al. 2007) and is thus in principle already set up for discourse processing. We integrate a pragmatic analysis of discourse particles via the XFR rewriting system (Crouch and King 2006, Crouch et al. 2008) and extend the AKR component to model CG based on Gunlogson.

2 Discourse Particles

German has a rich inventory of *discourse* or *modal* particles (von der Gabelentz 1891, Jacobs 1983, 1991, Abraham 1991, König 1997, Coniglio 2011). These particles have very subtle pragmatic content and contribute to the illocutionary force of an utterance (e.g., Karagjosova 2003, 2004, Bayer and Obenauer 2011). A subset of the German discourse particles establishes/invokes information with respect to a common ground (Zimmermann 2011), including *ja* 'yes', *doch* 'indeed' and *wohl* 'presumably'. We focus on these three, (2)–(4).[1]

(2) a. Das weißt du **ja**.
that know you **yes**
'You know that (already).'

b. Der Zug hat **ja** dann hinterher eine umso höhere Lageenergie.
the train has **yes** then afterwards an even higher potential energy
'The train will (of course) then afterwards have a higher potential energy.'
(from S21)

(3) a. Das ist **doch** klar.
that is **indeed** clear
'That is clear (as you ought to know).'

[1] A number of our examples are taken from the Stuttgart 21 mediation (S21) process, a public political discussion investigated as part of the *VisArgue* project on understanding political argumentation. Discourse particles were found to play a role in the framing of argumenation.

b. Es fährt **doch** ein TGV von Straßburg nach Stuttgart?
it drives **indeed** a TGV from Strasbourg to Stuttgart?
'There is a TGV from Strasbourg to Stuttgart, right?' (from S21)

(4) a. Er ist **wohl** nicht zu Hause.
he is **presumably** not at home
'Presumably he is not at home.'

b. Vielmehr ist es **wohl** genau so diskutiert worden.
rather is it **presumably** exactly so discussed become
'Rather, it has presumably been discussed exactly like that.' (from S21)

All three are very common German discourse particles and contribute expressive content that is associated with the CG. The particle *ja* invokes a fact as CG. *Doch* reactivates a fact that is assumed to already be part of the CG and *wohl* signals that speakers are either not entirely sure about their utterance, e.g. because they only heard of an event but did not witness it, or do not entirely trust the information conveyed by the proposition.

(5) [ja p] establishes or reconfirms the proposition p as part of the common ground (Lindner 1991, Kratzer 1999, Zimmermann 2011).

(6) [doch p] signals that the speaker assumes p not to be activated at the current state in the discourse, because the addressee may have temporarily forgotten about p or the addressee may consider p false (Lindner 1991, Karagjosova 2003, Zimmermann 2011).

(7) [wohl p] expresses a weakened commitment of the speaker to p. [wohl p] has been described as assume(x,p) (Kratzer, 1999, Zimmermann, 2011).

For example, in (2b), the speaker assumes that the hearer knows about the higher potential energy the train will have. In (3b) the speaker signals the addressee that they should already be aware of/believe that there is a TGV from Straßbourg to Stuttgart. And finally, *wohl* in (4b) conveys that the speaker presumes that a matter has been discussed in a certain way, but could not swear to it.

The particles are all highly ambiguous. For example *ja* also just means 'yes', the *doch* can also function as a subordinating conjunction and *wohl* could also be an adjective meaning 'fine, well'. The different usages can generally be disambiguated via syntactic or phonological cues or both. For example the discourse particle *ja* is usually used for acceptance and discourse structuring in its stressed variant and for manipulating certain beliefs in the common ground as illustrated above in its unstressed variant.[2]

[2]Steedman (2014) formulates a pragmatic account of English intonation that is also ultimately based on Stalnaker's CG. The illocutionary force of the German discourse is very similar to that described for English intonation (cf. Schubiger 1965, 1980). Intonation clearly also plays a pragmatic role in German and does interact with the discourse particles. This is an area in need of further work.

Bayer and Obenauer (2011) and Bayer and Trotzke (2015), who work within Minimalism, argue that German discourse particles are minor functional heads which do not project. They concentrate on *denn* 'then', *nur* 'only', *bloß* 'barely' and *schon* 'already'. These have focusing and scopal properties that particles like *ja* or *doch* do not have. As far as we can see, there is no clear (non theory internal) argument for functional head status. We therefore adopt the existing syntactic analysis of the German *ParGram* grammar (Dipper 2003), which treats discourse particles as syntactic adverbs.

3 Formal Pragmatics of Discourse

Stalnaker's (2002) CG model has served as a foundation for most if not all theories of discourse pragmatics. It assumes that all discourse participants in a discourse share a common set of presuppositions, namely, the shared beliefs that all speakers assume to be true for the sake of the conversation. While the main tenets of the model are widely accepted, the details continue to be subject to refinements and many ideas still await formalization.

The most widely accepted extension of Stalnaker's original approach is to use a set-theoretic model that describes the (public) beliefs of the discourse participants and the shared beliefs, i.e., the CG separately. This idea has been pursued by a number of researchers such as Ginzburg (1994), Gunlogson (2002) and Asher and Lascarides (2003). Our particular implementation draws heavily on Gunlogson's formal proposals for the CG.

We use Gunlogson's model since the formalization fits well with the computational and formal scaffolding already put in place by the AKR system within the XLE/XFR grammar development platform (Crouch et al., 2008). Gunlogson's model presupposes a comparatively small number of theorems to generate the different conditions that determine the roles of propositions in the discourse. The model is also well suited to addressing exactly the issues that arise with respect to discourse particles.

In keeping with the literature, we assume that a discourse model can be divided roughly into two parts. One is a collection of beliefs, commonly represented by propositions (section 3.2). The other is a set of rules that coordinates what is happening with these beliefs as the discourse progresses, i.e. a formal understanding of "discourse moves" (section 3.1). The next sections provide more details on this proposal and we illustrate our approach with respect to the discourse particle *doch* 'indeed' (section 3.3).

3.1 Discourse Moves

A formal model of discourse has to: a) collect beliefs that are relevant to the discourse (such as Stalnaker's CG mentioned above); b) record and encode *discourse moves* that describe how certain utterances manipulate the set of relevant

beliefs.[3] Some rather well understood discourse moves are, for example, assertion, acceptance and rejection (Walker, 1996). An important part in determining these discourse moves is covered by determining the logical consistency of the discourse. Logical consistency is calculated by taking a new proposition into account in conjunction with the illocutionary force, for example, as conveyed by discourse particles. The new proposition plus the illocutionary force are evaluated against the already existing belief sets and on the basis of this, assertation, acceptance or rejection is calculated.

Note that logical consistency between a new utterance and the CG is not a sufficient condition for acceptance. That is, if an utterance adds nothing new to the CG, it does not automatically convey acceptance of existing belief states in the CG. On the other hand, for rejection, logical inconsistency is a sufficient, but not a necessary condition. Thus, if an utterance is logically inconsistent with the CG, then it is to be classified as a rejection. However, there are other possibilities that also result in rejection which do not require logical inconsistency (Walker 1996).

Working within Segmented Discourse Representation Theory (SDRT), Lascarides and Asher (2009) argue that discourse moves such as accept must include the implicatures engendered by a given utterance. We agree with this and the German discourse particles can be seen as setting up implicatures. Lascarides and Asher's model includes discourse moves and ties these to a more sophisticated analysis of discourse in terms of rhetorical relations. Their model is formalized, but not implemented as part of an NLP application. Embedding the computational work presented in this paper into a more complex discourse model remains to be done.

3.2 Representation of the Common Ground

Our implementation is based on Gunlogson (2002). In this model a discourse participant is associated with a set of public beliefs, so called discourse commitments ($DC_{Speaker/Addressee}$). These have the following definition:

(8) Let DC_A and DC_B be sets of propositions representing the public beliefs of A and B, respectively, where:
   a. p is a public belief of A iff 'A believes p' is a mutual belief of A and B
   b. p is a public belief of B iff 'B believes p' is a mutual belief of A and B

These two sets together form the same set of propositions as in Stalnaker's notion of the CG. In other words, we can represent Stalnaker's CG by the ordered set $<DC_A, DC_B>$. We can translate the sets introduced above into sets of worlds which are called commitment sets ($cs_{Speaker/Addressee}$), as in (9).

(9) Let a discourse context C be $<cs_A, cs_B>$, where:
   a. $cs_A = \{w \in W$: A's public beliefs are all true of w$\}$

[3] From now on, for the sake of illustration, we assume a simple dialog between an speaker A and an addressee B when we talk about discourse.

b. $cs_B = \{w \in W$: B's public beliefs are all true of w$\}$
(Gunlogson, 2002)

The commitment sets allow us to classify propositions in terms of *propositional states*: (i) commitment, (ii) joint commitment (iii) unresolved, (iv) controversial. These propositional states feed into the determination of discourse moves and are crucial for identifying logical consistency within a discourse. The definitions as given by Gunlogson are in (10) and (11).

(10) a. A commitment is any proposition $p \in$ either $cs_{Speaker}$ or $cs_{Addressee}$
b. A joint commitment is a proposition $p \in cs_{Speaker}$ and $p \in cs_{Addressee}$

(11) Commitments may be:

a. unresolved, iff [neither p nor W-p $\in cs_{Speaker}$] and [$p \in cs_{Addressee}$][4]
b. controversial, iff W-p $\in cs_{Speaker/Addressee}$ and p is unresolved

A commitment can belong either to the beliefs of one of the discourse participants or to the beliefs shared by the discourse participants, the CG. Commitments can be either unresolved, which means that one of the discourse participants has committed to them, but not the other. A new commitment is typically unresolved, but it can also be controversial. In that case it contradicts the beliefs of the other discourse participant. A joint commitment on the other hand is a belief in the CG.

There are thus three main types of discourse moves: Those that manipulate the speaker's beliefs, those that modify the addressee's beliefs and those that modify the mutually shared beliefs. Furthermore, we can distinguish between assertive discourse moves and interrogative discourse moves.[5]

For an individual x a declarative has the prototypical shape in (12). This basically means that a declarative (a locution in Gunlogson) adds to a given set of beliefs (sets of worlds) those worlds where the declarative is true or rather removes those worlds where the content of the declarative is not true.

(12) $cs_X + S_{decl} = w \in cs_X$: the descriptive content of S_{decl} is true of w

The commitments of the speaker, the commitments of the addressee and the joint commitments can all be updated by a declarative.

3.3 Our Formalization

We illustrate the overall analysis via the discourse particle *doch* 'indeed'. This particle is ambiguous between a contradicting and a reminding reading, see (13) (repeated from (6)). These two situations presuppose different propositional states.

[4] Whereby W-p designates all the worlds in which p is not true.

[5] We concentrate on assertive discourse moves since a comprehensive description of interrogative discourse moves would go beyond the scope of this paper.

The reminding reading is only available in a context where the declarative is either a joint commitment or unresolved, but not controversial. The contradicting reading is only available when the propositional state is controversial.

(13) [doch p] signals that the speaker assumes p not to be activated at the current state in the discourse, because the addressee may have temporarily forgotten about p or the addressee may consider p false (Lindner 1991, Karagjosova 2003, Zimmermann 2011).

Thus, the particle *doch* can be used both for acceptance as well as rejection and serves as a good test case for our implementation. (14) summarizes the architecture underlying our computational implementation. The system stays true to Gunlogson's proposal although we work with slightly different sets of beliefs.

(14) a. SB (Speaker Belief) = {p: the speaker believes that p is true and the addressee is aware of that}

b. AB (Addressee Belief) = {p: the addressee believes that p is true and the speaker is aware of that}

c. $CG_{S\text{-}A}$ = {p: p is mutually believed by the speaker and the addressee}
$CG_{S\text{-}A}$ = {AB ∩ SB}

d. C (discourse context) = $_{S\text{-}A}$ = <SB,AB,$CG_{S\text{-}A}$>

In this formalization the beliefs of discourse participants are sets that contain facts about propositions p in the form "the speaker believes p". The speaker is assumed to have full access to the set of propositions in AB, the addressee's beliefs. The beliefs that are shared by both sets, i.e. the intersection, are stored in the $CG_{S\text{-}A}$, the CG (cf. also Caponigro and Sprouse (2007)).

Consider how this plays out with respect to the ambiguous discourse particle *doch* 'indeed'. Let S_{decl}+*doch* be a declarative with the descriptive content of S and the expressive content added by *doch*. There are two situations:

(15) a. iff $S_{decl+doch}$ is unresolved or a joint commitment in C + $S_{decl+doch}$, then C' = $CG_{S\text{-}A}$ + S_{decl}

b. iff $S_{decl+doch}$ is controversial in $S_{decl+doch}$ then C' = SB + S_{decl}

(15a) covers *doch* in its reactivating or reminding function. In this situation, the speaker asserts the descriptive content of the utterance to the mutually believed commitments, entering it into the CG. The rule in (15b) says that if *doch* occurs in a sentence whose descriptive content is controversial, then the reactivating reading is not available and the sentence only updates the beliefs of the speaker.[6]

[6]Some sentences with referential expressions cannot be analyzed this way. In that case the descriptive content of the referential needs to be added to the speaker's beliefs. In the following example the referential *Das* may refer to any sort of proposition made earlier in the dialog.

(16) Das glaube ich dir **doch**
that believe I you **indeed**
'I believe you (about this).'

4 Implementation

For the implementation we worked with the existing German ParGram grammar (Dipper, 2003), which already contained a syntactic treatment of German discourse particles.[7] The German grammar also already contained a basic semantic AKR component (Bobrow et al., 2007). The English AKR is fairly extensive and draws on a variety of lexical semantic resources that have been integrated, e.g. WordNet[8]. This remains to be done for the German grammar.

The AKR system is implemented via the XFR rule rewriting system (Crouch et al., 2008), which uses the f-structure as the basis for the construction of a semantic representation. The basic semantic representations are known as *Knowledge Representations*, or KRs, which are then further processed into *Abstract Knowledge Representations* or AKRs that are suitable for inferencing and *Entailment and Contradiction Detection* (ECD) over several facts/utterances.

The AKR for (17) is as in (18). AKRs consist of two components, the *conceptual structure* and the *contextual structure*. The conceptual structure provides facts that are necessary for reasoning such as cardinality and onotological information and it partially represents the predicate-argument structure. Pointers to lexical semantic information are encoded in square brackets. This provides the system a good underlying basis for drawing inferences based on ontological information.

(17) Boris singt.
Boris sings
'Boris is singing.'

(18)

Conceptual Structure:	Contextual Structure:
subconcept(singen:7,[sing-1])	context_head(t, singen:7)
subconcept(Boris:1,[person-1]	top_context(t)
role(Agent,singen:7,Boris:1)	instantiable(singen:7,t)
role(cardinality_restriction,Boris:1,sg)	instantiable(Boris:1,t)

The information that is contained in the conceptual structure remains unchanged by our system. In the next two sections, we describe the modifications of the grammar and the AKR system that work with the contextual structure in order to provide a computational modeling of the CG as described above.

4.1 Syntactic Extensions

By default, the ParGram grammars parse text one sentence at a time. However, the grammars can be modified so as to allow for the processing of multiple sentences within one parse. We modified the German grammar to allow the parsing of small dialogs as in (19), where the speaker is identified at the beginning.

[7] A version of the German grammar can be used interactively via the INESS website: http://clarino.uib.no/iness/xle-web.

[8] https://wordnet.princeton.edu/

(19) a. Anna: Boris singt.
Boris sings
'Anna: Boris is singing.'

b. Clara: Daniel tanzt.
Daniel dances
'Clara: Daniel is dancing.'

As an illustration, the parse of (19) is given in Figures 1 and 2. The modified grammar uses the category ROOT as the top-level node. This expands to ROOTPunct and ROOTmulti. ROOTpunct directly governs a sentence and ROOTmulti recursively connects all ROOTpunct sentences. The +SB indicates a sentence boundary and is necessary for the proper tokenization of the utterances.

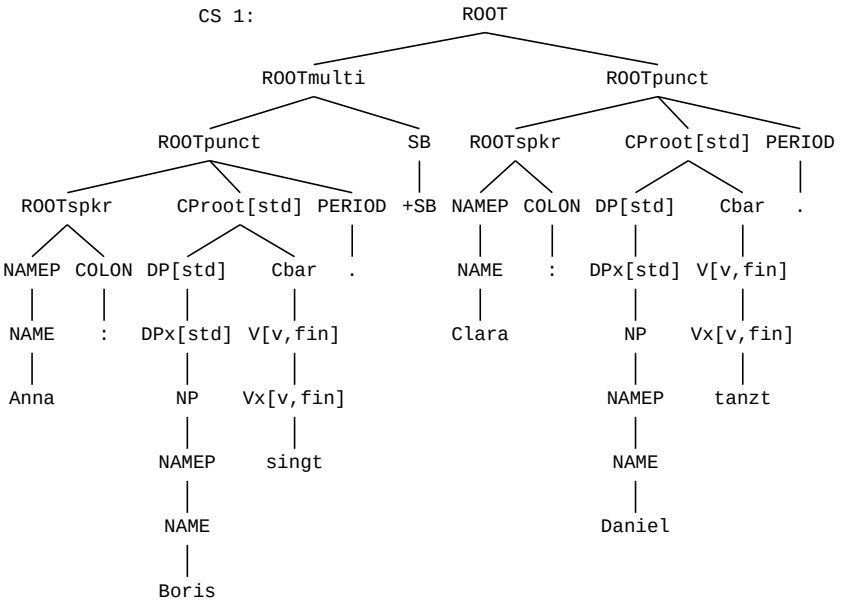


Figure 1: C-structure of the two sentences in (19).

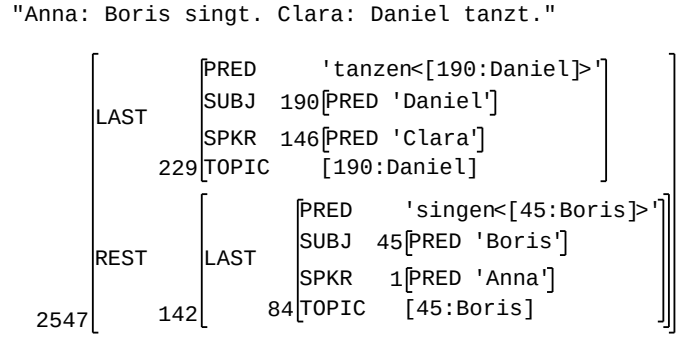


Figure 2: F-structure of the discourse in (19)

The category ROOTspkr registers the speaker information. This speaker information is also represented at f-structure under the attributed SPKR. Information

about which utterance came first, i.e., the temporal structure of the discoure, is represented at f-structure via a LAST vs. REST partition.[9] The most recent sentence in a discourse is encoded in the outermost f-structure while the first sentence of the discourse is encoded in the innermost f-structure. Although it may seem counterintuitive, this f-structure representation best fits with the existing treatment of contextual information within AKR, as discussed in the next section.[10]

4.2 Pragmatic Extensions

In our implementation, each sentence in a parse is connected to a root node, i.e., `ROOT` in single sentence parses or `ROOTPunct` in multi sentence parses. Speaker information also associates with root nodes. The system requires that each sentence necessarily is directly or indirectly governed by at most one speaker. To explain the importance of this connection it is first necessary to understand the process of rewriting a syntactic structure into a semantic structure.

F-structures as in Figure 2 serve as the basis for the semantic analysis. XLE accesses the f-structure as a list of predicate-argument facts that represent the attribute-value pairs and computes a basic KR. In a KR all facts that make up the semantics of a sentence are connected to some context. These contexts are what are used to model phenomena such as epistemic modality. Contexts are introduced at the root-level of an f-structure. In Crouch and King's (2006) version of the XFR semantics, matrix sentences associate with top-level contexts, i.e., contexts that contain information that is assumed to be true in the actual world. In comparison, embedded contexts with the label `ctx:predicate` are invoked by `COMP`s and `XCOMP`s. Embedded contexts are linked to top level contexts. As a result all contexts in a sentence are rooted in a top-level context. In other words, all contexts are associated with root nodes. Embedded contexts are connected to the top-level context via context-lifting relations, such as `veridical`, induced for example by factive verbs like *know*, or `crel` induced, for example, by *believe*. These relations determine whether the propositional content of a `COMP` or `XCOMP` is true at the top-level context or not.

In our implementation both speaker information and top-level contexts are associated with root nodes. We connect the speaker information directly to the top-level contexts so that facts are associated with `<speaker,context>` pairs instead of only contexts. The semantics then make the facts relative to the speaker's top-level context of the sentence. Unwinding the f-structure information in order to do this via the Prolog logic underlying the construction of the AKRs is most easily achieved via the Rest-Last ordering of utterances shown in Figure 2.

As the revised contextual structure in (20) for (19) shows, the major changes

[9] This was originally introduced in ParGram to allow for the parsing of fragmentary input.

[10] In this paper we assume that one continuous list of utterances associated with one speaker is summed up into one contribution. As pointed out by an anonymous reviewer, it may be necessary to split up the contributions into single sentences and analyze them separately to account better for the scope of certain discourse moves. We leave this for future work.

to the AKR involve changes with respect to instantiability facts. These statements now tell us which pieces of information are believed to be true in which contexts relative to which speakers.

(20) Contextual Structure:
context([Anna]:t)
context([Clara]:t)
context([Clara,Anna]:t)
top_context([Clara,Anna]:t)
instantiable(Boris:7,[Clara,Anna]:t)
instantiable(Daniel:27,[Clara,Anna]:t)
instantiable(singen:13,[Anna]:t)
instantiable(tanzen:34,[Clara]:t)

Other Structure:
discourse_participants([Clara,Anna])

In the original AKR facts are mapped to contexts that roughly represent the actual and possible worlds. In our modified AKR the contextual facts are provided with another dimension, the dimension of speaker commitment. That is, these facts are mapped to those worlds which the speaker associated with them believes to be true. As a result, the instantiability facts tell us exactly who believes what.[11]

The remainder of the AKR construction makes sure that these facts are only true in the right conceptual environment (e.g. predicate argument structure) via the conceptual structure. Overall, the changes to the existing AKR component were rather minimal since many of the prerequisites for a modeling of the CG in discourses were already in place.

In conclusion, out implementation allows us to model the set-theoretic assumptions of Gunlogson's discourse models with AKRs. Facts can be instantiable with respect to (i) the speaker, (ii) the addressee, (iii) the speaker and the addressee and the context they are associated with respectively.

4.3 Modeling the Common Ground — Discourse Moves

The pragmatic parser relies heavily on the Entailment and Contradiction Detection (ECD) that is part of the XLE/XFR platform and that was originally used as part of a Q&A system that the original AKR was designed for. In the original system, the ECD operates on a query that is matched to a set of stored facts and detects whether a stored fact matches the query. The match could be a direct match or be the result of a chain of inferencing. For example given the stored information that

[11] We assume entities denoted by proper names to be presupposed. That is, each discourse participant accepts the existence of an individual denoted by a proper name and can judge utterances made about that individual. As a reviewer points, out, this assumption is open to debate. We leave this for future work. Note also, that while the industrial strength version of the English grammar did incorporate a component for anaphora resolution, we do not have access to such anaphora resolution for the German grammar. Thus, we make the simplifying assumption for now that a proper name always denotes the same entity. This is the case for both speaker information and proper names in the discourse. When treating larger diaologs, we can employ unique indices instead.

a boy rode a bicycle, the system can infer that the answer to the query *Did a child ride a bicycle?* is "yes" via the information that a boy is a subconcept of *child*.

In our system, we crucially rely on the ECD to determine the type of discourse move a given utterance represents. This classification in turn steers the dynamic update of the CG. As in the original Q&A task, our implementation aligns a set of stored facts, namely the existing belief states, with a query. This "query" is the AKR representing the newest contribution to a dialog. The ECD checks whether it is possible to remove all query facts via a system that unifies passage and query. If unification is not possible, the system checks whether query and passage logically contradict each other and if so flags the fact that is responsible as contradictory.

The ECD thus returns one of three possible answers: `YES` (corresponding alignment) in case of entailment, `NO` in case of contradiction and `UNKOWN` otherwise. Each new utterance is assigned one of the three propositional states shown in (21), depending on the result of the ECD query. The ECD result triggers a specific set of rules for rewriting the CG as illustrated by prototypical rules in (21). The antecedent of the rules in (21) illustrates the shape of the facts in the query (i.e. the new contribution), while the consequent denotes the shape of the fact in the AKR of the complete discourse. The rules are simplified for the sake of illustration.[12]

(21) commitment – default: `instantiable(fact,[speaker:t)])`

joint commitment:	ECD returns `YES`
	`instantiable(fact,[speaker]),`
	`instantiable(fact,[addressee]) ==>`
	`instantiable(fact,[speaker,`
	`addressee]:t)`

unresolved:	ECD returns `UNKNOWN`
	`instantiable(fact,[speaker:t]) ==>`
	`instantiable(fact,[speaker:t])`

controversial:	ECD returns `NO`
	`instantiable(fact,[speaker:t]) ==>`
	`instantiable(controversial,fact,`
	`[speaker:t])`

Based on the propositional states the system can make hypotheses about the intended discourse moves, i.e., whether the speaker intends to accept or reject previously uttered information. Acceptance leads to a simple update of the belief states. The speaker accepts a proposition as joint commitment. Rejection is more complex. In principle a new update to the CG that is controversial with respect to the previous CG will most likely indicate a rejection move. For now, the system always concludes a rejection move if it detects a controversial propositional

[12]The rules are also applied in the same fashion to uninstantiable facts. Also note that the unresolved rule actually does nothing and is listed here only for purposes of illustration. In the implementation the system simply does nothing in this case.

state.[13] The facts that belong to the new update are than flagged as such. The flag is necessary to guarantee the integrity of the AKR, since otherwise we would have contradicting facts in the AKR representing the discourse structure. This in turn would lead to obvious problems for the ECD.

Discourse particles and similar phenomena further affect how information is treated by the rewrite rules. Recall that *ja* 'yes' and *doch* 'indeed' both may add facts to the CG. Rules ensuring this pragmatic inferencing triggered by discourse particles are applied after the ECD check. The rules belong to the set of rules that combine the previous CG with the information added by the new discourse contribution introduced in (21). This means we first check the "at-issue" conent of the new proposition and then in a further step parse its expressive content as part of updating the CG. The following sequence of rules are applied for each new utterance to be added to the CG: (i) load the existing CG-AKR; (ii) create a "query" AKR from the new discourse contribution; (iii) run ECD rules; (iv) run rule set for combining CG-AKR and the query AKR based on ECD information and other pragmatic cues such as those triggered by discourse particles.

4.4 Pragmatic Parsing of Discourse Particles

In this section, we discuss our treatment of *ja* 'yes', *doch* 'indeed' and *wohl* 'presumably'. They each present an different test case. The discourse particle *ja* represents a case in which the right reading of the particle needs to be identified by a non-pragmatic module. The particle *doch* must be disambiguated as part of the pragmatic parsing and *wohl* illustrates how the system deals with representing different layers of commitment.

For the sake of illustration, we always assume direct temporal precedence of the relevant content. In other words, as before, we assume very small segments of discourse consisting of an utterance and a response to that utterance.

4.4.1 The Particle *ja*

The role of *ja* in discourse is two-fold if we follow the analysis of Zimmermann (2011). Either, it is a discourse-structuring particle that marks acceptance as well as rejection (stressed *ja*), or it is a discourse particle that marks that an utterance should be in the CG (unstressed *ja*, e.g. in (2)). In other words: an utterance made

[13] However, the time span between the update and the contradicting fact in the previous discourse needs to be taken into consideration. The current implementation is not time span sensitive since we have only implemented analyses of very short discourse pieces. The system architecture has three states: *direct temporal precedence* referring to the information in the CG that was changed by the previous update; *temporal precedence* for all information that lies beyond that point in the past and *disconnected precedence*. The latter is information that happened in a previous discourse segment which is temporally independent from the current discourse. The beliefs that fall under the notion *temporally independent* roughly include everything that is not part of the currently analyzed discourse, e.g. common general knowledge and possibly information from a previous discourse. I.e. all facts that the discourse participants mutually believe to be true before the start of the current dialog.

with unstressed *ja* either adds its content to the CG or repeats something that is already in the CG. These two readings can be disambiguated primarily by phonological cues (stressed vs. unstressed). Given the first option, we can disambiguate between acceptance and rejection via a logical consistency check and thus determine the discourse move the speaker intended.

The repetition reading is only available when the conveyed content is not controversial. The expressive content that is added to the meaning of a sentence by the particle *ja* asserts the semantic content of the sentence to the CG. This means that if the conveyed content so far has the status of being unresolved or is already registered as a commitment, it is updated to a joint commitment in the belief states. In this sense, the stressed *ja* and the unstressed *ja* overlap in meaning. An assertive discourse move can still be rejected later on, as illustrated in (22).[14]

(22) Context of dialog: Conversation between two people who talk about the animal they are looking at in the petting zoo.

Speaker:	Das ist **ja** ein Kaninchen.
	That is **yes** a rabbit
	'This is a rabbit (as I want you to know).'
Addressee:	Ich glaube, das ist ein Hase.
	I think that is a hare
	'I think this is a hare.'

The system captures the intuition that both the stressed and the unstressed *ja* add something to update the CG and that *ja* can also be used for the purpose of rejection. The various readings of *ja* must be identified via a combination of syntactic and phonological cues.

4.4.2 The Particle *doch*

The discourse particle *doch* has two different readings. A rejection reading and a (re)activation reading (Zimmermann 2011, Karagjosova 2003). Both readings can be inferred by the state of the *doch*-proposition in the CG. Generally, the rejection reading occurs whenever there is logical inconsistency with respect to the CG or the beliefs of the addressee. This normally occurs as direct consequence of the last discourse moves. Thus, direct temporal precedence, as we presuppose in this section, plays a factor in analyzing *doch* as indicator for rejection.[15]

The activation reading of *doch* applies if *doch* + proposition is so far unresolved in the CG. This means the ECD returns neither a contradiction nor a confirmation. The rewrite rules that modify the CG then "activate" the proposition and make it part of the CG. We illustrate how this works in our system via the short dialog

[14]There is also a third reading of *ja*, where *ja* is a sign of being impressed or surprised. This reading is in principle also available in (22).

[15]*doch* may also appear as the only word in an utterance. In this case, it conveys rejection. When *ja* is uttered by itself, in contrast, it conveys acceptance.

in (23), where Anna makes an assertion that is not directly contradicted by Clara. Rather, she uses *doch* to (re)activate the information in the CG that Boris also dances. The contextual information in the AKR in (24) records the belief states of the individual participants and of the CG (which in this case is what both Anna and Clara believe).

(23) a. Anna: Boris singt. / Boris sings
'Anna: Boris sings.'

b. Clara: Boris tanzt doch. / Boris dances indeed
'Clara: Boris dances (as you should know).'

(24) a. **Common Ground: Anna: Boris singt.**

Contextual Structure:
context([Anna]:t)
context_head([Anna]:t,singen:13)
top_context([Anna]:t)
instantiable(Boris:7,[Anna]:t)
instantiable(singen:13,[Anna]:t)

Other Structure:
discourse_participants([Anna])

b. **New contribution: Clara: Boris tanzt doch.**

Contextual Structure:
context([Clara]:t)
context_head([Clara]:t,tanzen:13)
top_context([Clara]:t)
instantiable(Boris:7,[Clara]:t)
instantiable(tanzen:13,[Clara]:t)

Other Structure:
discourse_participants([Clara])

c. **Result of ECD:**
proposition state: new information;
discourse move: assertion

d. **Update/Rewriting in accordance with ECD:**
Anna: Boris singt. Clara: Boris tanzt doch.

Contextual Structure:
context([Anna]:t)
context([Clara]:t)
context([Clara,Anna]:t)
top_context([Clara,Anna]:t)
instantiable(Boris:7,[Clara,Anna]:t)
instantiable(singen:13,[Anna]:t)
instantiable(tanzen:33,[Clara,Anna]:t)

Other Structure:
discourse_participants([Clara,Anna])

Information in the CG is matched with a new contribution. The Query Match detects no contradiction and the result of the ECD analysis is that the new contribution is classified as the discourse move of assertion due to the contribution of the discourse particle *doch*. This triggers an update of the CG as shown in (24d).

4.4.3 The Particle *wohl*

The discourse particle *wohl* works differently than *ja* 'yes' and *doch* 'indeed' in that *wohl* expresses a weakened commitment to a proposition. In our system, we allow for two forms of weakened commitment: (1) temporally weakened, if there is a large temporal distance between the addition of the proposition to the CG and the utterance *wohl* + p; (ii) semantically weakened if the information is connected to an embedded context instead of the top level context.

Recall that the AKR already includes a modeling of embedded propositions in terms of allowing for non-veridical embedded propositions for verbs such as *believe* (vs. *know*). This is achieved via the `context` predicate in the AKR in conjunction with context-lifting relations (`veridical, crel`), see section 4.4. We adapt the existing AKR treatment of embedded propositions to *wohl*.

All of the pragmatic parsing steps also apply to *wohl*. The existing (if any) CG-AKR is loaded. A "query" is formulated from the new discourse contribution and the "at-issue" content is checked for contradictions. The ECD component is then applied and new information is integrated into the CG as appropriate.

Consider the dialog in (25), where the addressee offers a conjecture as to what the cause of the loud music might be (Boris is dancing). The *wohl* conveys that the addressee is making a conjecture, not stating a fact.

(25) Anna: Very loud music is playing next door.
Clara: Boris tanzt **wohl**.
Boris dances **presumably**
'Presumably, Boris is dancing.'

The parsing of the discourse particle *wohl* induces an operation similar to that of *ja* and *doch*, but it updates an embedded context instead of the top-level context, as *ja* and *doch* do. This models the existing analysis of *wohl* + proposition as `assume(proposition)` (Zimmermann, 2011).

As shown in (26), the embedded context and the top-level context are connected by the context relation that indicates embedding of the context `ctx` under the top-level context `t`. The context `ctx` represents a different level of commitment of p, namely `assume(speaker,p)`, compared to the context `t` which expresses something like `know(speaker,p)`.

(26) **New contribution: Clara: Boris tanzt wohl.**
Contextual Structure:
context([Clara]:t)
context([Clara]:ctx(tanzen:32))
context_head([Clara]:t,wohl:18)
top_context([Clara]:t)
context_relation([Clara]:t,[Clara]:ctx(tanzen:32),wohl:18))
instantiable(Boris:26,[Clara]:ctx(tanzen:32))
instantiable(Boris:26,[Clara]:t)
instantiable(tanzen:32,[Clara]:ctx(tanzen:32))

The system treats the relation asserted in the AKR as a `crel` relation. This means the speaker has no credible source that the information related to the embedded context is true in the actual world. As part of the CG update after the ECD step, this context is updated for all of the discourse participants with the facts comprised in *wohl* + proposition. The effect is that the proposition modified by *wohl* enters the CG of all the discourse participants, but with the status of an assumption.

Another condition for this analysis is that *wohl* + proposition has the status of being unresolved so far in the beliefs of at least the speaker. If this is not the case, for example, if the proposition represents a commitment of one of the discourse participants, then that person's belief set does not change.

4.5 Summary of Implementation

We have shown how to represent belief states of discourse participants via AKRs and how to keep track of individual and CG knowledge. We have also shown how discourse particles contribute information pertinent to the CG and in terms of the discourse moves of a speaker.

The subtle differences between *ja*, *doch* and *wohl* are captured by our analysis. As can be seen in (27), there is some overlap in the meaning of these discourse particles, but our analysis takes into account the slight nuances by which they alter the pragmatic content of an utterance.

(27)

unstressed *ja*	if for *ja* + p, p is either unresolved or a joint commitment $\longrightarrow$ p is added to the CG.
doch	if for *doch* + p, p is either unresolved or a weakened commitment $\longrightarrow$ p is added to the CG.
wohl	if for *wohl* + p, p is either unresolved or a weakened commitment $\longrightarrow$ p is added as weakened commitment to the CG.

The analysis is implemented via XFR rewrite rules. For each contribution to the CG the system calculates the AKR of the new utterance and determines its propositional state with respect to the AKR representing the CG. This is achieved by using the ECD system.

The system also calculates discourse moves and flags them if necessary. The flagging is necessary, for example, for discourse moves of a controversial nature that may lead to rejection.[16] Controversial moves in a discourse participant's AKR are rewritten as soon as a new proposition made by the discourse participant contradicts it and this proposition is logically consistent with the CG. This means that the discourse participant revises their own beliefs and accepts the beliefs of the

[16] It is also necessary for questions, for example, which do not add anything to the CG. There may be other discourse moves that require special flags but this is left for future work. This would mainly concern discourse moves that do not serve to structure or to update the CG.

other discourse participant. As a result this requires the ECD to be sensitive to the propositional state and flags such as `controversial`. Ultimately, no information about discourse moves surfaces in the AKR. The AKR only encodes information about the beliefs of the discourse participants.

5 Conclusion and Outlook

Our analysis and implementation is capable of capturing the finely nuanced meanings of the discourse particles and uses these to organize the beliefs in the discourse structure. Our paper presents a formal pragmatic approach for analyzing discourse particles. Specifically, we employed a discourse model based on set-theoretic notions devised by Gunlogson (2002). For this discourse model we described means to identify some basic discourse moves in accordance with Walker (1996). Finally, we established the connection between Gunlogson's model, discourse moves and discourse particles. In other words, we elaborated on how discourse particles operate on the CG in interaction with discourse moves.

We illustrated the workings of our pragmatic parser with respect to the German discourse particles *ja*, *doch* and *wohl*. Our analysis has a strong orientation towards natural language processing tasks and we implemented our analysis by means of the XLE/XFR grammar development platform. We used the German ParGram grammar as a basis, augmenting it to deal with multi-sentence discourses and extending the existing AKR semantic component to register speaker belief states, calculate discourse moves and keep track of information in the CG. We found that the AKR was already set up well for our task as there are parallels between the Q&A PARC bridge system the AKR was originally designed for (Bobrow et al., 2007) and the entailment and contradiction detection that is needed for modelling CG. The overall system is modular and can be extended in several different ways to apply to broader research questions than the ones pursued here.

References

Abraham, Werner. 1991. Discourse particles in German: How does their illocutive force come about? In W. Abraham, editor, *Discourse Particles: Descriptive and Theoretical Investigations on the Logical, Syntactic and Pragmatic Properties of Discourse Particles in German*, Pages 203–252. Amsterdam: John Benjamins.

Asher, Nicholas and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press.

Bayer, Josef and Hans-Georg Obenauer. 2011. Discourse particles, clause structure, and question types. *The Linguistic Review* 28:449–491.

Bayer, Josef and Andreas Trotzke. 2015. The derivation and interpretation of left peripheral particles. In J. Bayer, R. Hinterhölzl, and A. Trotzke, editors., *Discourse-oriented Syntax*. Amsterdam: John Benjamins. To Appear.

Bobrow, Daniel G., Bob Cheslow, Cleo Condoravdi, Lauri Karttunen, Tracy Holloway King, Rowan Nairn, Valeria de Paiva, Charlotte Price, and Annie Zaenen. 2007. PARC's bridge and question answering system. In T. H. King and E. M. Bender, editors., *Proceedings of the GEAF 2007 Workshop. CSLI Studies in Computational Linguistics Online*, Pages 46–66. CSLI Publications.

Caponigro, Ivano and Jon Sprouse. 2007. Rhetorical Questions as Questions. In *Proceedings of Sinn und Bedeutung*, volume 11, Pages 121–133.

Condoravdi, Cleo, Dick Crouch, Valeria De Paiva, Reinhard Stolle, and Daniel G. Bobrow. 2003. Entailment, Intensionality and Text Understanding. In *Proceedings of the HLT-NAACL 2003 Workshop on Text Meaning*, volume 9, Pages 38–45. Association for Computational Linguistics.

Coniglio, Marco. 2011. *Die Syntax der deutschen Modalpartikeln: Ihre Distribution und Lizensierung in Haupt- und Nebensätzen*, volume 73 of *Studia grammatica*. Berlin: Akademie Verlag.

Crouch, Dick, Mary Dalrymple, Ron Kaplan, Tracy King, John Maxwell, and Paula Newman. 2008. *XLE Documentation*. Palo Alto Research Center.

Crouch, Dick and Tracy Holloway King. 2006. Semantics via F-Structure Rewriting. In M. Butt and T. H. King, editors., *Proceedings of the LFG06 Conference*, Pages 145–165.

Dipper, Stefanie. 2003. *Implementing and Documenting Large-Scale Grammars – German LFG*. Ph.D. thesis, IMS, University of Stuttgart.

Ginzburg, Jonathan. 1994. An Update Semantics for Dialogue. In H. Bunt et al., editors., *Proceedings of the first International Workshop on Computational Semantics*.

Gunlogson, Christine. 2002. Declarative Questions. In B. Jackson, editor, *Proceedings of Semantics and Linguistic Theory XII*, Pages 124–143. Ithaca, NY: CLC Publications.

Jacobs, Joachim. 1983. *Fokus und Skalen: Zur Syntax und Semantik der Gradpartikeln im Deutschen*. Tübingen: Niemeyer.

Jacobs, Joachim. 1991. On the semantics of modal particles. In W. Abraham, editor, *Discourse Particles*, Pages 141–162. Amsterdam: John Benjamins.

Karagjosova, Elena. 2003. Modal Particles and the Common Ground: Meaning and Functions of German *ja*, *doch*, *eben/halt* and *auch*. In P. Kühnlein, H. Rieser, and H. Zeevat, editors., *Perspectives on Dialogue in the New Millennium*, volume 114, Pages 335–349. Amsterdam: John Benjamins.

Karagjosova, Elena. 2004. *The Meaning and Function of German Modal Particles*. Ph.D. thesis, Universität des Saarlandes.

König, Ekkehard. 1997. Zur Bedeutung von Modalpartikeln im Deutschen: Ein Neuansatz im Rahmen der Relevanztheorie. *Germanistische Linguistik* 136:57–75.

Kratzer, Angelika. 1999. Beyond "oops" and "ouch": how descriptive and expressive meaning interact. In *Cornell Conference on Theories of Context Dependency*, volume 26.

Lascarides, Alex and Nicholas Asher. 2009. Agreement, disputes and commitments in dialogue. *Journal of Semantics* 26(2):109–158.

Lindner, Katrin. 1991. 'Wir sind ja doch alte Bekannte' — the use of German *ja* and *doch* as modal particles. In W. Abraham, editor, *Discourse Particles: Descriptive and theoretical investigations on the logical, syntactic and pragmatic properties of discourse particles in German*, volume 12, Pages 163–201. Amsterdam: John Benjamins.

Schubiger, Maria. 1965. English intonation and German modal particles: a comparative study. *Phonetica* 12:65–84. reprinted in **?**, 175-93.

Schubiger, Maria. 1980. English intonation and German modal particles ii: a comparative study. In L. Waugh and C. van Schooneveld, editors., *The Melody of Language*, Pages 279–298. Baltimore, MD: University Park Press.

Stalnaker, Robert. 2002. Common Ground. *Linguistics and Philosophy* 25(5/6):701–721.

Steedman, Mark. 2014. The surface compositional semantics of English intonation. Draft: a revised version is to appear in *Language* 90, 2-57.

von der Gabelentz, Georg. 1891. *Die Sprachwissenschaft, ihre Aufgaben und Methoden*. Tübingen: Narr.

Walker, Marilyn A. 1996. Inferring acceptance and rejection in dialogue by default rules of inference. *Language and Speech* 39(2-3):265–304.

Zarrieß, Sina. 2009a. Developing German semantics on the basis of parallel LFG grammars. In *Proceedings of the 2009 Workshop on Grammar Engineering Across frameworks (GEAF 2009), Singapore*, Pages 10–18. Association for Computational Linguistics.

Zarrieß, Sina. 2009b. From parallel syntax towards parallel semantics: Porting an English LFG-based semantics to German. In *Proceedings of the Biennal GSCL Conference 2009, Potsdam*, Pages 159–169. Tübingen: Narr.

Zimmermann, Malte. 2011. Discourse Particles. In K. v. Heusinger, C. Maienborn, and P. Portner, editors., *Semantics. (= Handbücher zur Sprach- und Kommunikationswissenschaft HSK 33.2)*, volume 2, Pages 2011–2038. Berlin: De Gruyter.