

# NOMINAL ARGUMENTS IN DUTCH AND WYSIWYG LFG

Annie Zaenen  
Xerox©Parc, CSLI©Stanford

## Introduction

Syntacticians rather frequently exhibit a type of behavior that seems to greatly puzzle practitioners of other crafts: they regret bitterly that the tools that they have at their disposal to ply their trade are too powerful. The generally accepted reasoning goes as follows: mechanism  $x$  would allow a grammar  $y$ , leading to linguistic behavior  $z$ , surely we never see  $z$ , so we should not have the full power of  $x$ . We need a constraint. But the constraints proposed on the basis of this reasoning rarely achieve their aim: the ingenuity displayed by the linguistic community greatly exceeds its desire and/or ability to formalize its proposals in a watertight way. Any proposed constraint can always be circumvented in a way to allow  $z$ , should  $z$  contrary to prior assumptions occur, as is often the case. In the face of this result syntacticians exhibit another puzzling kind of behavior: they are not unduly troubled, certainly not moved to mend their ways either towards better formalizations or towards abandoning the constraint trade. They often even do not revoke the constraint for which the motivation seems to have disappeared. To make sense of this behavior I hypothesize that the formulation of constraints plays actually a slightly different role from the proclaimed one: it is not meant to exclude any overt state of affairs but to guide hypothesis formation. To take an example that will play some role in what follows, in several frameworks one finds a constraint with the effect that every verb needs to have a subject. In the light of the countless number of sentence types without subjects in most languages of the world, the constraint is either immediately falsified or has to be supplemented by the hypothesis that some subjects are phonologically null. From an empirical point of view the claim that all verbs have a subject but that in not too well understood circumstances some subjects are allowed to remain silent encodes as much lack of knowledge as the claim that in certain ill understood circumstances some verbs may remain subjectless (or for that matter the claim that in certain ill understood circumstances, verbs require subjects). As a guiding principle for further research the different proposals have, however, a quite different effect: the first one leads to the further hypothesis that the 'occurrence' of these null subjects will be regulated by some possibly rather menial principles of morphology and the like and allow the theory of deep syntactic phenomena to proceed as if indeed all verbs had subjects, whereas the latter formulation does not lead to such a claim.

If the real function of constraints is indeed to guide hypothesis formation, a first principle should be that the constraints are not built into the notation/formalization of the linguistic theory itself: if one formulates one's theory in such a way that no alternatives to it can be entertained within the language of the theory itself, one will not be able to compare different ways of approaching a problem and hence be condemned from the beginning on to be totally right or totally wrong. Given the odds that the latter is the case, this does not seem to be a desirably state to be in. So I take it as a first desideratum that constraints be things that can be added or left out without an immediate effect on the overall formalism. This deprives us of course of one possible source of constraints: formalizations.

An obvious other possible source is language acquisition: as has often been observed, different languages are learned at more or less the same rate by people with vastly different levels of intelligence. What this tells us about possible grammars, however, depends on what one assumes can be plausibly taken to be innate and is further colored by the related issue of how much variation among the grammars of different languages one thinks it is reasonable to hypothesize. I will take an unpopular stance on both these issues and assume that a great deal of language acquisition is indeed language learning and that most of what is innate and plays a role in language acquisition is not specific to language. I also take it that 'universal grammar' is still an hypothesis to be proven instead of being a given but I will follow most current schools of grammatical thought in taking the position that the best way to show whether the hypothesis is right or wrong is by trying to isolate some principles of UG. As I do not take the UG hypothesis as obviously right, however, I do not want to get pushed into the position of having to assume a universal principle to maintain UG and then having to declare that principle innate because it couldn't possibly be acquired in any other way.

My view on UG and innateness does not provide much guidance in the choice between general linguistic tools e.g. transformational or functionally extended PS grammars but it leads to different evaluation of the principles or constraints that are proposed to guide hypothesis formation than the more currentview does. One of the issues it bears most directly on seems to me to be that of empty categories: if we assume that most linguistic generalizations cannot be grounded in innate principles it would be prudent to propose only inaudible or invisible elements for which the evidence is relatively direct, otherwise a theory of language acquisition with the postulated characteristics will be difficult to develop. Motivated by these concerns I propose in the next section some realistic constraints on lexical functional grammars and in the remainder of the paper I will show

how these principles can be maintained in an account of some phenomena that have been handled previously in ways that violate them.

### Some background assumptions.

As there are no up to date published introductions to LFG, I will summarize the overall view of the framework that I am taking as my point of departure. In LFG, the linguistic representation of a sentence (or text) is a multi dimensional object consisting of several 'projections' (footnote: "levels, refs to Fenstad et al and Halvorsen....). I will briefly characterize the different projections and the correspondence rules between them.

#### **Projections:**

*a. the thematic projection:* this projection contains information about individual predicates and their arguments. At this point I am assuming a simple list of thematic roles and a partial order on them.(footnote: Jackendoff) The list and the ordering are both rather familiar (see Kiparsky, forthcoming, Bresnan and Kanerva, 1988) ; in this paper only the following will play a role

(1a) agent < cause < goal, experiencer < theme ...

It is assumed that the notions used in this projection can eventually be grounded in semantic notions. In as far as this is true it is desirable from a realistic point of view to describe as many generalizations as possible in these terms rather than in terms that are more remote from non linguistic experience.

*b. the syntactic projection:* this projection has the structure familiar from earlier work in LFG. Whereas the thematic projection is a representation containing individual predicates, the syntactic projection is a representation of the sentence as a whole. It contains not only information about the arguments of predicates but also about the integration of adjuncts and the like. The structure is hierarchical, allowing multi dominance. The substantive ingredients of this projection are grammatical roles (GR's) (footnote: I will talk about roles instead of functions: given the way the framework is set up technically all these different projections have functions but, when talking about things like subjects and objects one tends to talk one time about a function and another time about the value of the function, just to avoid that confusion I will avoid the word). The change with respect to earlier work is that, as discussed in Levin (1986, 1988) and Bresnan and Kanerva (1988), the OBJ2 role is replaced by a series of OBJ<sub>n</sub> roles. For

the purposes of this paper this is a purely notational change. I will also assume a special unit on this level, the 'nucleus' which is composed of a predicate and its arguments (check with Joan) and an ordering among the main GF's as proposed in Zaenen, Maling, and Thráinsson (1985), given in (1b):

(1b) SUBJ < OBJ < OBJth < ? OBL.

Given the *parti pris* in favor of concreteness and against the innateness of purely linguistic concepts expressed above, the adoption of this projection might come a bit as a surprise: whereas notions like thematic roles can hopefully be grounded in 'cognitive structure' and discourse notions in a theory of communication and action, the syntactic projection is purely linguistic. As will be clear at the end of the paper, I hope, there is little of substance to notions like subject and object in my approach, what is at issue is that there is a hierarchy that is different from the thematic and from the discourse hierarchy; nothing is said about it being not derivable.

*c. the discourse projection:* here linguistic entities are classified according to their discourse role. Work on this projection is in its preliminary stages. Following Bresnan and Mchombo (19876), a classification of constituents in focussed and non-focussed is presupposed here. Some hypotheses about the organization of this projection in Dutch will be given in section 2.

*d. the PS projection:* it is characterized by an ordering which corresponds to the temporal ordering of the words in the uttered sentence, an hierarchical tree structure and category labels.

### Correspondence Rules.

The correspondence rules between the thematic and the functional projection have been discussed in Levin (1986, 1988) and Bresnan and Kanerva (1988). The main idea is that thematic roles are incompletely specified with respect to the GF's they are linked to. General and language or rule specific principles allow the full specification. The exact specifications used in this paper will be discussed in section 2. Direct mappings from the thematic projection to the PS projection are also allowed and discussed in section ...and...The correspondence rules between the PS projection and the functional projection are most extensively discussed by Bresnan (1982) who proposes some principles regulating this mapping depending of the configurational or non configurational encoding of GF's. I will come back to this issue in section 2.3.2. The mapping between the discourse projection on the one

hand and the functional projection and the PS projection on the other has received very little attention. What is proposed in this paper can be found in sections .... and ...

The architecture of a linguistic theory here is very different from that found in serial transformational frameworks: wellformedness constraints in all the different projections contribute to a sentence being grammatical or ungrammatical. The interactions are not limited through assumption that there can be only one 'feeding' relation between each pair of projections. Such an architecture allows for interactions that in a transformational framework would be either ruled out or be represented in a very unelegant way. Given that there are a number of cases of the latter (find refs), this is a desirable result. The framework imposes its own general constraints because each projection has to be defined and has no recursive mechanism to generate more than one representation.

Not only the technical setup is different; the underlying philosophy is too: the arguments of predicates (to limit ourselves to the linguistic elements that will be our focus of attention in this paper) are represented at all projections, their behavior will be a function of their thematic role, their GF and discourse function and their syntactic category and nothing more.

## **1. Proposed constraints.**

The constraints that I will discuss are imposed on the functional projection. They are the following two, in keeping with the realistic approach outlined above:

- (2) The identification constraint on GR's (IC):
  - a grammatical role has to be identifiable.
  - There are two ways to identify a GF:
    - its phonological form,
    - its thematic role.

To give this constraint empirical content we have to make the technical requirement that every GR has either a PRED or a FORM value and that FORM values can only be introduced directly in lexical entries. GR can also be indirectly identified through f-control chains. On these, we need the following constraint:

- (3) Constraint on f-control:
  - A GR cannot be linked by f-control to another GF of the same predicate.

The second principle is intended to rule out equations like (SUBJ)=(OBJ), which, if allowed, would immediately void the first constraint of all empirical content. It is reminiscent of the constraint on thematic roles imposed in Williams (1987).

The first constraint is intended to insure that all phonologically null material is either assigned a thematic role by a predicate of which it is an argument or indirectly through chain formation. The constraint has the effect that all such chains have to be thematically linked to a PRED or be phonologically realized.

In what follows I will give an analysis of Dutch constructions which have been previously analysed as having an inaudible dummy subject, in a way that respects these constraints. Most of these involve agentless predicates. The first class contains some experiencer constructions (section 2) which I will compare to a type of psychological predicate with different characteristics (section 3). I will then extend this analysis to the passives of ditransitives (section 4). Finally I will turn to sentences with indefinite subjects, agents or no agents, and to impersonal passives. In all sections word order constraints are relevant. It will be argued that Dutch word order is the result of the interplay between constraints in the thematic projections, the syntactic projection and the discourse projection. To discuss the discourse factors in any detail, sentences with indefinite subjects need to be looked at.

This is only done in section 5. In the earlier sections I will keep the influence of discourse factors constant by limiting the discussion of word order to definite non pronominal arguments, i.e. arguments that on the scale from presupposed to focussed tend to fall somewhere in the middle and are both of the same type. A further limitation on the scope of the inquiry is that only nominal arguments will be considered, by this I mean bare NP arguments, PP's remain outside of the scope of this paper.

Even if I am successful, this does of course not establish that the constraints can be maintained in all circumstances but it will illustrate the mechanisms that LFG has at its disposal to obviate the need for inaudible dummies in favor of a more concrete characterization of the properties of verbal arguments.

## **2. A potential problem: experiencer verbs**

Dutch has several classes of predicates with an experiencer argument. The class that I will discuss first can be characterized by the following two properties: the past tense auxiliary is in general *zijn* (to be) (footnote: *bevalen*, *smaken* are exceptions), say something about the

predictability of aux selection) and it has no passive forms. Some verbs falling into that class are given in (4):

- (4) bevallen, ontglippen, ontvallen, ontgaan, smaken,...  
please, escape, elude, taste,...

As the attempts in (5) illustrate, they cannot be passivized either as impersonal passives or as personal passives:

- (5a)\* ...de boeken worden (door de jongen) bevallen.  
...the books (pl) are (by the boys) pleased.  
'..the books are liked by the boy.'
- (5b)\* ...de jongen wordt (door de boeken) bevallen.  
...the boy (sg) is (sg) (by the books) pleased.  
'..the boy is pleased by the books.'
- (5c)\* ...er wordt (door de boeken) de jongen bevallen.  
..there is (sg) (by the books) the boy pleased.
- (5d)\* ..er worden/wordt (door de jongen) de boeken bevallen.  
...there are(pl)/is(sg) (by the boy) the books pleased.

There is another class of mainly psychological predicates in Dutch which has different characteristics. It will be discussed in section 3. Some examples are given in (6):

- (6) ergeren, vervelen, treffen,...  
annoy, bore, be-noticed,...

### 2.1. Some superficial characteristics of experiencer verbs

The arguments of the predicates in (4), exemplified in (7), behave in some ways differently from the arguments of simple transitive verbs illustrated in (8):

- (7) ...Jan de boeken zijn bevallen.  
...John the books (pl) are (pl) pleased.  
'..the books have pleased John.'
- (8) ...Jan de boeken heeft gelezen.  
...John the books (pl) has (sg) read.  
'..John has read the books.'

As is customary, the examples are given as they would occur in embedded clauses: in declarative main clauses any major constituent

can be topicalized, distorting the basic word order facts. I will refer to the order exemplified in embedded clauses as the order in the 'middle field'. (footnote: direct questions, origin of middle field, extension to include subject)

Superficially, the construction exemplified in (7) provides contradictory information about the grammatical functions of its arguments: if we take case marking and agreement to be indicators of what the grammatical role of a Dutch NP is, we have to conclude that in sentences like (7) and (9), *de boeken* (the books) and *hij* (he) respectively are the subject because the verb agrees with them and in (9), *hij* is clearly in the nominative (footnote about case marking). I will refer to this argument as the 'theme' or the 'agreeing argument'.

- (9) ...hij hen is bevallen.  
...he (nom) them (nnom) is (sg) pleased.  
'..he has pleased them.'

If on the other hand one assumes that word order is a good indicator of grammatical role in Dutch and that the subject is to be identified as the first (nominal) argument in a clause, it seems that the non nominative argument, which I will refer to as the 'experiencer' or the 'non agreeing argument', can be the subject: in (8) the experiencer argument is in first position. (10) shows that both orders are possible with full NP's (footnote on pronouns)

- (10) ...de jongen de boeken zijn bevallen.  
...the boy (sg) the books (pl) are (pl) pleased.  
'..the books have pleased the boy'.

In general we do not find such a discrepancy between word order and case marking or agreement as tests for grammatical functions with definite NP's: in a sentence like (12), only the version in which the NP that triggers agreement is also in first position is completely grammatical.

- (11a) ...de jongen de boeken heeft gelezen.  
...the boy (sg) the books (pl) has (sg) read.  
'..the boy has read the books.'

- (11b)?\*...de               boeken de jongen heeft gelezen  
...the books (pl) the boy (sg) has (sg) read.  
'..the boy has read the books.'

The 'coding properties' (footnote Keenan) for canonical subjects do not seem to pick out either argument of a verb like *bevallen* (to please) in



an unambiguous way. Universal principles about case marking or word order do not decide the matter in a straightforward way either: there are languages with non nominative subjects (see e.g. Zaenen, Maling and Thrainsson, 1985, for a summary of the arguments for such subjects in Icelandic) and there are languages in which word order is determined by other factors than grammatical functions (see Uszkoreit, 1985, for a detailed discussion of the factors that influence German word order in the middle field).

Given this apparent conflict in the superficial tests for subjecthood with these predicates, there are three types of hypotheses that come to mind. It could be the case that either argument can be a subject: whichever argument is in first position is the subject. This is the analysis proposed in Levin (1986). A second possibility is to take the view that the nominative is always the subject and to maintain the case marking/agreement test for subjects. This is the path that I will pursue here. A third possibility is that in sentences in which the theme follows the experiencer argument, there is no overt subject. The word order constraints are respected, albeit vacuously and agreement needs to be accounted for by further assumptions. This is the analysis of predicates with an experiencer argument as 'impersonal inversion constructions' in Relational Grammar terms. Within that framework it has not been proposed for Dutch as far as I know but it would be the only consistent analysis for the Dutch data within the RG framework. Similar ideas have been put forward within a GB framework by Hoekstra (1984). I will argue against the first hypothesis because it introduces the need for unmotivated further assumptions. The third hypothesis in its LFG incarnation is ruled out by the constraints that I formulated in section 1. To make these constraints plausible then, I have to provide an account based on hypothesis 2 that is at least as adequate as an account along the lines of 3 would be.

## 2.2. The proposed analysis.

Under the analysis to be defended, the predicates under consideration have the following thematic representation:

(12) bevallen: experiencer < theme.

and the following assignment of grammatical functions:

(13) experiencer = OBJth.  
theme = SUBJ.

A rudimentary theory of the mapping of thematic roles to GR's is given in Bresnan and Kanerva (1988). Grammatical roles are seen as

decomposable into two features: +/- r (for restricted) and -/+ o (for objective). These features give us the following four classes of GF's:

- |      |       |    |    |
|------|-------|----|----|
| (14) | SUBJ  | -r | -o |
|      | OBJ   | -r | +o |
|      | OBJth | +r | +o |
|      | OBL   | +r | -o |

Intuitively, the two features reflect the observation that some thematic roles, e.g. themes are more often encoded as subjects or objects than as obliques or secondary objects, whereas others, e.g. agents are more often encoded as either subjects or obliques and not as objects of any kind. This is made explicit by the following intrinsic specifications (we give here only the assumptions we start off with, others will be proposed later) :

- (15) Intrinsic specifications:  
 agent: -o (universal) (not relevant in this section)  
 theme: -r (universal)

Other specifications are introduced by rules or by defaults. The defaults are given in (16):

- (16) Defaults:  
 the highest role (on the hierarchy given in section ....) is marked -r, the next one +o and the third one +r.

Bresnan and Kanerva (1988) also adopt a universal subject condition. In my view there is no empirical evidence for such a proposal and in languages like German and Dutch it can only be maintained at the cost of violating the identification constraint proposed in 1 (see section 9). Instead of it, I propose the following weaker condition which as far as I know is consistent with the facts of Chichewa as well (footnote about Alsina's new analysis) as of those of languages like German and Dutch to the defaults:

- (17) Assign the highest thematic role to the subject role.

(16) and (17) together can most likely be replaced by the following more intuitive condition, which I will call the RV (revised defaults):

- (17') Default: thematic roles are, in their hierarchical order (as given in (1a) mapped to the highest GR (as given in (1b)) compatible with their specifications.(footnote)

The defaults as their name implies come only into play when no other specifications would conflict with their use.

A supplementary assumption is that in Dutch and German the experiencer argument of the verbs under consideration is lexically case marked, as dative in German and as nonnominative in Dutch. This case marking leads to the addition of the +r feature to that argument by a general principle given in (18):

- (18) In languages in which -r marked arguments are assigned case according to their GR (i.e. by non configurational encoding), lexically marked case assignment implies +r assignment.

I will refer to this principle as the UC (unique case assignment principle); it is meant to say that when a language identifies the semantically unrestricted grammatical roles (in the sense of Bresnan (1982) through case marking, it can't use, possible conflicting, lexical case marking to mark these same arguments.

These assumptions allow the following mappings:

- (19) Experiencer: +r and +/- o, i.e. OBJth (by UC and RD)  
Theme: -r, -o (by the intrinsic marking and the RD)

Bresnan and Kanerva propose that passive morphology adds the +r marking to the highest thematic role. This could be interpreted as allowing for passive morphology to be added vacuously in the case of the predicates at hand: +r would be assigned to the experiencer argument which already has this specification. I will assume that such vacuous application of morphologically conditioned rules is not allowed and that the lack of passivization of these predicates is accounted for in this way.(footnote: impersonal passives)

As said I will ignore prepositional phrases. Most of them are obliques in Dutch and their behavior is quite different from that of the nominal arguments that are investigated here. The proposed feature distinction between arguments given here does not treat nominal arguments as a natural class. This is to my mind a shortcoming of the proposal but one that will not be addressed here.

As far as the PS rules are concerned, I see no need to assume the existence of a VP in the middle field in Dutch. Ignoring 'verb-raising', the basic rule of S expansion is:

$$(20) \quad S = \{XP^*, V\}$$

Saying that the lexical case marking of the experiencer entails its +r marking is equivalent to saying that in Dutch, like in German, the encoding of the major GR's is non configurational in the sense of Bresnan (1982), i.e. these functions are not directly encoded via PS positions but through functional annotations that follow the general scheme given in (21):

$$(21) \quad \begin{array}{l} \text{XP} \\ (\text{GR}) \\ (\text{Feature})=f \end{array}$$

where the specific function selected is constrained by the value of a feature like case. The kind of attributes that may replace the GF variable will be discussed further in section 4.2.1.

The following ordering constraints are assumed to hold in the c-structure:

$$(22) \quad \text{agent or SUBJ} < \text{experiencer} < \text{theme}$$

(interpretation needs to be discussed). Generalizations of this ordering will be discussed later.

The discourse projection for Dutch contains the ordering constraint given in (23), taken from Verhaegen (1982) and the principle in (24), taken from Bennis (1986):

$$(23) \quad \text{Presupposed} < \text{Focussed}$$

$$(24) \quad \text{there is always at least one presupposed element in the middle field.}$$

This concludes the overview of my analysis. In the next section, I give arguments in favor of the analysis under which only the theme argument of *bevallen* has subject properties in Dutch (subsection 2.3.1) and discuss the evidence that might be taken to argue in favor of an alternative postulating that the experiencer argument can also be the subject (subsection 2.3.2.). I will conclude that the simplest analysis treats only the theme as a subject.