

**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 *SIE* – 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, *SIE* can be used in more than one function at the same time; there is also the issue of haplology of *SIE* in verb chains and of sharing it under coordination – the proposed analysis captures these phenomena. The paper discusses the status of *SIE*, 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 *SIE*, 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 *SIE* in Polish and briefly discusses their properties. § 3 shows that *SIE* may bear more than one function in relation to the same predicate. § 4 presents the phenomenon of haplology of *SIE* in verb chains, while § 5 discusses sharing of *SIE* under coordination of verbal predicates.

§ 6 is devoted to the issue of the status of *SIE*: 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 *SIE* – as a marker. However, this issue is discussed further in § 7, which shows potential problems with such an analysis of *SIE*. Finally, § 8 concludes the paper.

It is worth noting that Polish *SIE* 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 *SIE* and haplology of *SIE* 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 *SIE* (which provides detailed information about the type of *SIE* involved, providing an explicit account of ambiguous readings with *SIE*); it also includes an analysis of sharing *SIE* under coordination, which is not included in Kupść 2000.

¹All links provided in this paper were accessed on 20 December 2015.

²Sharing of *SIE*, 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.’

³Though the word *nago* is translated as ‘nude’ in (4), it is an adverb – as indicated in glosses.

⁴Morphosyntactic 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 SIE_ę 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 SIE_ę are a lexical requirement of the given predicate: these include reflexive (cf. (1)), reciprocal (cf. (2)) and inherent (cf. (3)) SIE_ę.

By contrast, impersonal SIE_ę 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 SIE_ę 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 SIE_ę contributes to the semantics – it is the case with all types of SIE_ę discussed above except for inherent SIE_ę, which is required syntactically but it does not make any contribution to the semantics. By contrast, the remaining types of SIE_ę make such a contribution: the reflexive SIE_ę marks that the respective predicate is reflexive, the reciprocal SIE_ę marks that it is reciprocal, influencing the semantic interpretation, while the impersonal SIE_ę marks that the predicate is impersonal.

2.4 Agreement constraints

When the impersonal SIE_ę 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 SIE 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 SIE

It is possible for SIE to have more than one function with respect to the same predicate at the same time: in both examples below SIE is impersonal (as in (4)), apart from being inherent in (16) (as in (3)) and reflexive in (17) (as in (1)). Multifunctionality of SIE is represented in glosses by joining appropriate basic types of SIE 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 SIE as dependents of the same predicate, so the use of multifunctional SIE 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 SIE. (16) has a reading where SIE 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.wirtualnemediia.pl/teologowie-moralisci-apelujadomedioiw-t9375.html>

⁶<https://odjehani.com.pl/Thread-Pierwsze-golenie?page=2>

(17) in turn has two alternative readings: one where SIĒ is only reflexive and the subject is implicit (as discussed for (20)), shown in (21), and the other where SIĒ is only impersonal (with an implicit subject that blocks the use of a lexical one, see § 2.4) and the predicate GOLIC̄ 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 Haplogy of SIĒ 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 SIĒ, 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 SIĒ, as in (23).⁷ Such sharing of SIĒ in verb chains is represented in glosses by joining the types of SIĒ 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 SIĒ (as shown in (24)), but there is only one textual instance of SIĒ, which structurally belongs to the higher predicate, namely ZDARZYĆ, as demonstrated in (25), which is ungrammatical with SIĒ 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 haplogy of SIĒ is also possible when, unlike in (23), the respective verbs in the verb chain require different types of SIĒ:

⁷CZY, glossed as Q in (23)–(25), is the Polish *yes/no* question particle.

⁸<http://www.druga-strona.pl/oobe-podroz-pozza-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 SIE, while its infinitival complement, namely GOLIĆ ‘shave’, takes the reflexive SIE – as shown in (27). Similarly, in (28), the main verb CHODZIĆ ‘walk’ forms a construction with the impersonal SIE, while the embedded predicate GAPIĆ ‘stare’ requires the inherent SIE – as demonstrated in (29).

Finally, as mentioned when discussing multifunctionality of SIE, sentences with SIE may be ambiguous depending on the interpretation of SIE. This is the case in (26), which has an alternative reading presented in (30) where haplology is not involved – SIE is only a dependent of the main verb (which takes inherent SIE) 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 SIE under coordination

Like other dependents, SIE 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 SIE 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 co-ordinated verbs – see (31)–(32), where CAŁOWAĆ ‘kiss (each other)’ and PRZYTU-
 LAĆ ‘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 Ibrahimović
 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 ad-
 mires 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).

¹¹http://www.pudelek.pl/arttykul/63056/rozenek_i_majdan_razem_w_mcdonaldzie_calowali_sie_i_przytulali/

$$(38) \left[\begin{array}{l} \text{PRED 'ADMIRE'} \langle \boxed{1} \rangle \\ \text{SUBJ } \boxed{1} \left[\begin{array}{l} \text{PRED 'ZLATAN'} \\ \text{CASE NOM} \end{array} \right] \end{array} \right]$$

$$(39) \left[\begin{array}{l} \text{PRED 'ADMIRE'} \langle \boxed{1}, \boxed{2} \rangle \\ \text{SUBJ } \boxed{1} \left[\begin{array}{l} \text{PRED 'ZLATAN'} \\ \text{CASE NOM} \end{array} \right] \\ \text{OBJ } \boxed{2} \left[\begin{array}{l} \text{PRED 'SELF'} \\ \text{CASE ACC} \end{array} \right] \end{array} \right]$$

The formalisation of the proposed analysis of SIE 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 SIE discussed in § 3.

6.1 Templates used in lexical entries

The template called in the lexical entry of SIE is provided in (40):

$$(40) \text{ SIE} \equiv (\uparrow \text{ SIE PRESENT}) = + \wedge [\text{ SIE-LEX-IMPRS} \vee \text{ SIE-IMPRS-ONLY}]$$

$$(41) \text{ SIE-LEX-IMPRS} \equiv \text{ SIE-LEX} \wedge (\text{ SIE-IMPRS})$$

$$(42) \text{ SIE-IMPRS-ONLY} \equiv \text{ SIE-IMPRS} \wedge \neg \text{ SIE-LEX}$$

$$(43) \text{ SIE-LEX} \equiv (\uparrow \text{ SIE } \{\text{ INH|REFL|RECIP}\}) =_c +$$

$$(44) \text{ SIE-IMPRS} \equiv (\uparrow \text{ SIE IMPRS}) = + \wedge (\uparrow \text{ SUBJ}) = \%S \wedge (\%S \text{ PRED}) = \text{'PRO'} \\ \wedge (\%S \text{ PERS}) = 3 \wedge (\%S \text{ NUM}) = \text{SG} \wedge (\%S \text{ GEND}) = \text{N}$$

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 SIE, 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 SIE 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 SIE (reflexive, reciprocal or inherent), which accounts for lexically required SIE, 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 SIE 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 SIE, while the second conjunct is a negated (\neg) call to the SIE-LEX template – its result is precluding SIE from being of any of the types defined in (43), which includes inherent, reflective and reciprocal SIE. The effect of these constraints is that SIE 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 SIE – they contain calls to the template defined in (45), where the value of the parameter TYPE corresponds to the type of required SIE (INH, REFL or RECIP; impersonal SIE 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 SIE, while the second one ensures that SIE 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 SIE discussed in § 4. Sample lexical entries for verbs used in examples (1)–(3) are provided below:

- (46) *golić* V (\uparrow PRED) = ‘SHAVE<(\uparrow SUBJ)>’ \wedge @(SIE-TYPE REFL)
 (47) *całować* V (\uparrow PRED) = ‘KISS<(\uparrow SUBJ)>’ \wedge @(SIE-TYPE RECIP)
 (48) *śmiać* V (\uparrow PRED) = ‘LAUGH<(\uparrow SUBJ)>’ \wedge @(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) SIE, respectively. Due to the adopted analysis of SIE 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.

$$(49) \left[\begin{array}{l} \text{PRED 'SHAVE'} \langle \boxed{1} \rangle \\ \text{SUBJ } \boxed{1} \left[\begin{array}{l} \text{PRED 'JACEK'} \\ \text{CASE NOM} \\ \text{NUM SG} \end{array} \right] \\ \text{SIE } \left[\begin{array}{l} \text{REFL } + \\ \text{PRESENT } + \end{array} \right] \end{array} \right] \quad (50) \left[\begin{array}{l} \text{PRED 'KISS'} \langle \boxed{1} \rangle \\ \text{SUBJ } \boxed{1} \left[\begin{array}{l} \text{PRED 'PRO'} \\ \text{CASE NOM} \\ \text{NUM PL} \end{array} \right] \\ \text{SIE } \left[\begin{array}{l} \text{RECIP } + \\ \text{PRESENT } + \end{array} \right] \end{array} \right]$$

The representation of the remaining two types of SIE is not controversial: (51) corresponds to (3), which features inherent (INH) SIE, while (52) provides a sim-

plified representation of (4), which contains an impersonal (IMPRS) SIE.

$$(51) \left[\begin{array}{l} \text{PRED 'LAUGH}(\underline{1})' \\ \text{SUBJ } \underline{1} \left[\begin{array}{l} \text{PRED 'PRO'} \\ \text{CASE NOM} \\ \text{NUM SG} \\ \text{PERS 2} \end{array} \right] \\ \text{SIE } \left[\begin{array}{l} \text{INH } + \\ \text{PRESENT } + \end{array} \right] \end{array} \right]$$

$$(52) \left[\begin{array}{l} \text{PRED 'WALK}(\underline{1})' \\ \text{SUBJ } \underline{1} \left[\begin{array}{l} \text{PRED 'PRO'} \\ \text{GEND N} \\ \text{NUM SG} \\ \text{PERS 3} \end{array} \right] \\ \text{SIE } \left[\begin{array}{l} \text{IMPRS } + \\ \text{PRESENT } + \end{array} \right] \end{array} \right]$$

6.3 Multifunctional SIE

Let us now consider some structures produced by this analysis for sentences that involve multifunctional SIE discussed in § 3 – the f-structures in (53)–(54) provide a representation of simplified versions of (16) (*śmiało się z czerwonych*) and (17) (*golito się żyletkami*).

$$(53) \left[\begin{array}{l} \text{PRED 'LAUGH}(\underline{1},\underline{2})' \\ \text{SUBJ } \underline{1} \left[\text{PRED 'PRO'} \right] \\ \text{OBL } \underline{2} \left[\begin{array}{l} \text{PRED 'RED'} \\ \text{PFORM Z} \\ \text{CASE GEN} \end{array} \right] \\ \text{SIE } \left[\begin{array}{l} \text{INH } + \\ \text{IMPRS } + \\ \text{PRESENT } + \end{array} \right] \end{array} \right]$$

$$(54) \left[\begin{array}{l} \text{PRED 'SHAVE}(\underline{1})' \\ \text{SUBJ } \underline{1} \left[\text{PRED 'PRO'} \right] \\ \text{ADJ } \left\{ \begin{array}{l} \text{PRED 'RAZOR'} \\ \text{CASE INST} \end{array} \right\} \\ \text{SIE } \left[\begin{array}{l} \text{REFL } + \\ \text{IMPRS } + \\ \text{PRESENT } + \end{array} \right] \end{array} \right]$$

In (16) SIE is at the same time inherent (specified lexically by LAUGH) and impersonal (constructionally), while in (17) SIE is both reflexive (as specified by SHAVE) and impersonal.

Such multifunctional specifications are the result of the interaction of lexical entries of verbs requiring SIE lexically and the lexical entry of SIE (see (40)), which contains a call to the template SIE-LEX-IMPRS defined in (41) that can optionally introduce the specification of the impersonal SIE (given in (44)) if it satisfies the condition formalised in (43), namely that the given instance of SIE 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 SIE

Under the current analysis, SIE 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 SIE is satisfied as a result of haplology do not have this attribute, which makes it possible to identify instances of SIE 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:

$$(55) \left[\begin{array}{l} \text{PRED} \quad \text{'FEAR'} \langle \underline{1}, \underline{2} \rangle \\ \text{SUBJ} \quad \underline{1} \left[\text{PRED} \quad \text{'MANY'} \right] \\ \text{XCOMP} \quad \underline{2} \left[\begin{array}{l} \text{PRED} \quad \text{'SHAVE'} \langle \underline{1} \rangle \\ \text{SIE} \quad \left[\text{REFL} \quad + \right] \end{array} \right] \\ \text{SIE} \quad \left[\begin{array}{l} \text{INH} \quad + \\ \text{PRESENT} \quad + \end{array} \right] \end{array} \right]$$

The higher predicate, FEAR, to which SIE belongs structurally, contains two attributes inside SIE: PRESENT, contributed by the lexical instance of SIE, and INH, which corresponds to the type of SIE 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 SIE which is required by SHAVE – since this predicate does not have a lexical SIE as its dependent but it uses haplogy 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 SIE (using a defining equation in (45)), it is possible to handle haplogy where different types of SIE are involved, as in (26) discussed above, where the main verb takes the inherent SIE, while its infinitival complement takes the reflexive SIE. Similarly, in (28) SIE belonging to the main verb is impersonal (constructional, not required lexically as in (26)), while the infinitival complement takes the inherent SIE.

Finally, such an analysis of haplogy ensures that the requirement of an appropriate type of SIE 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 SIE (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.’

$$(57) \left[\begin{array}{l} \text{PRED} \quad \text{'FEAR'} \langle \underline{1}, \underline{2} \rangle \\ \text{SUBJ} \quad \underline{1} \left[\text{PRED} \quad \text{'PRO'} \right] \\ \text{XCOMP} \quad \underline{2} \left[\begin{array}{l} \text{PRED} \quad \text{'TRY'} \langle \underline{3} \rangle \underline{1} \\ \text{SUBJ} \quad \underline{1} \\ \text{XCOMP} \quad \underline{3} \left[\begin{array}{l} \text{PRED} \quad \text{'SMILE'} \langle \underline{1} \rangle \\ \text{SIE} \quad \left[\text{INH} \quad + \right] \end{array} \right] \end{array} \right] \\ \text{SIE} \quad \left[\begin{array}{l} \text{INH} \quad + \\ \text{PRESENT} \quad + \end{array} \right] \end{array} \right]$$

- (58) Bał się próbować (*się) uśmiechnąć się.
 feared.SG.M1 INH try.INF INH smile.INF INH

6.5 Sharing SIE under coordination

The last remaining issue is sharing of SIE under coordination. While examples where SIE 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 SIE 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 SIE:

- (59) IP_{top} → IP DEP Conj IP
 $\downarrow \in \uparrow$ $\uparrow = \downarrow$ $\uparrow = \downarrow$ $\downarrow \in \uparrow$
- (60) DEP → { ARG | MOD | RM }
 $(\uparrow \text{GF}) = \downarrow$ $\downarrow \in (\uparrow \text{ADJUNCT})$ $\uparrow = \downarrow$

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 SIE:

- (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) Zakleiała 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 SIE of the same type (reciprocal), while in (34) they share SIE 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, SIE. Secondly, since SIE, 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 SIE), 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 SIE_i 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 SIE_i would make it impossible to account for sentences such as (34), where conjuncts require different basic types of SIE_i – since particular verbs call (45) to specify the type of required SIE_i, two different calls would result in a multifunctional specification of SIE_i, 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 SIE_i, namely reciprocal SIE_i:

$$(63) \left[\left[\begin{array}{l} \text{PRED 'KISS'} \langle \underline{1} \rangle \\ \text{SUBJ } \underline{1} \left[\begin{array}{l} \text{PRED 'PRO'} \\ \text{CASE NOM} \\ \text{NUM PL} \end{array} \right] \\ \text{SIE} \left[\begin{array}{l} \text{RECIP } + \\ \text{PRESENT } + \end{array} \right] \end{array} \right] , \left[\begin{array}{l} \text{PRED 'HUG'} \langle \underline{1} \rangle \\ \text{SUBJ } \underline{1} \\ \text{SIE} \left[\begin{array}{l} \text{RECIP } + \\ \text{PRESENT } + \end{array} \right] \end{array} \right] \right] \right] \\ \text{COORD-FORM AND}$$

The f-structure in (64) corresponds to (34), where coordinated verbs take different types of SIE_i (inherent vs. reflexive). Accounting for the fact that the coordinated predicates in (34) require two different basic types of SIE_i is possible due to the fact that the call to the template SIE defined in (40) (placed in the lexical entry of SIE_i) introduces the defining equation (\uparrow 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 SIE_i – as discussed in § 6.1, it is the verb that specifies the type of SIE_i using the defining equation in (45).

$$(64) \left[\left[\begin{array}{l} \text{PRED 'LAUGH'} \langle \underline{1} \rangle \\ \text{SUBJ } \underline{1} \left[\begin{array}{l} \text{PRED 'PRO'} \\ \text{CASE NOM} \\ \text{GEND M1} \\ \text{NUM PL} \\ \text{PERS 3} \end{array} \right] \\ \text{SIE} \left[\begin{array}{l} \text{INH } + \\ \text{PRESENT } + \end{array} \right] \end{array} \right] , \left[\begin{array}{l} \text{PRED 'KNOCK'} \langle \underline{1}, \underline{2} \rangle \\ \text{SUBJ } \underline{1} \\ \text{OBL } \underline{2} \left[\begin{array}{l} \text{PRED 'HEAD'} \\ \text{PFORM W} \\ \text{CASE ACC} \\ \text{NUM PL} \end{array} \right] \\ \text{SIE} \left[\begin{array}{l} \text{REFL } + \\ \text{PRESENT } + \end{array} \right] \end{array} \right] \right] \right] \\ \text{COORD-FORM AND}$$

The proposed analysis also accounts for the impossibility of sharing SIE_i under

¹³As shown in (63) and (64), the implicit subject is shared by the coordinated verbs.

coordination where the predicate in one conjunct would lexically require SIE_E (inherent, reflexive or reciprocal), while with the other SIE_E 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 SIE_E 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 SIE_E, is defined – its second conjunct is a disjunction of two template calls: (41) and (42). The latter, (42), introduces the exclusively impersonal SIE_E, which fails to satisfy the requirements of POTYKAĆ ‘stumble’ in (65). By contrast, the former, (41), introduces lexically required SIE_E (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 SIE_E – 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 SIE_E to relevant predicates in examples such as the following one, where only some conjuncts require SIE_E:

- (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 SIE_E, 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 SIE_E, but this would not only distort the semantics of the relevant predicates as SIE_E would be impersonal in all conjuncts, but it would also conflict with the inherent SIE_E specification introduced lexically using (45) by the edge conjuncts (BAĆ and ŚMIAĆ), as (42) requires SIE_E 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 SIE_E as a marker:

¹⁴Two 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 SIE as a marker, this should not be possible when SIE 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 SIE 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
 ciocie 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 SIE is treated as a case-bearing argument, but not when it is always claimed to be a marker.

Further evidence against treating SIE 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) SIE 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, SIE_Ė 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 SIE_Ė behaves in the same way as SIEBIE, which, according to Polish binding rules, typically¹⁶ 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)
$$\left[\begin{array}{l} \text{PRED} \quad \text{'ORDER}(\underline{1},\underline{2},\underline{3})\text{' } \\ \text{SUBJ} \quad \underline{1} \left[\text{PRED} \text{'JAN'} \right] \\ \text{OBJ}_{\theta} \quad \underline{2} \left[\text{PRED} \text{'PIOTR'} \right] \\ \text{XCOMP} \quad \underline{3} \left[\begin{array}{l} \text{PRED} \text{'BUY}(\underline{2},\underline{4},\underline{5})\text{' } \\ \text{SUBJ} \quad \underline{2} \\ \text{OBJ} \quad \underline{4} \left[\text{PRED} \text{'BOOK'} \right] \\ \text{OBJ}_{\theta} \quad \underline{5} \left[\text{PRED} \text{'SELF'} \right] \end{array} \right] \end{array} \right]$$

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 SIE_Ė: 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 SIE_Ė 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 SIE_Ė (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 SIE_Ė is bound by the local subject, the hairdresser.

¹⁶SIEBIE 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.’

¹⁷The glosses and free translation in (73) were added for the purposes of this paper.

$$(76) \left[\begin{array}{l} \text{PRED} \quad \text{'ORDER}(\underline{1},\underline{2},\underline{3})\text{' } \\ \text{SUBJ} \quad \underline{1} \left[\text{PRED} \quad \text{'FATHER'} \right] \\ \text{OBJ}_{\theta} \quad \underline{2} \left[\text{PRED} \quad \text{'HAIRDRESSER'} \right] \\ \text{XCOMP} \quad \underline{3} \left[\begin{array}{l} \text{PRED} \quad \text{'SHAVE}(\underline{2})\text{' } \\ \text{SUBJ} \quad \underline{2} \\ \text{SIE} \quad \left[\begin{array}{l} \text{REFL} \quad + \\ \text{PRESENT} \quad + \end{array} \right] \end{array} \right] \end{array} \right]$$

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 reflexive (and possibly reciprocal) SIE as also having an incarnation as an argument, which would lead to a split analysis of SIE. This, however, requires further research.

8 Conclusion

This paper offered an LFG analysis of SIE – the Polish so-called reflexive marker – showing its different uses, providing an account of a number of related phenomena, which include multifunctionality of SIE, where it has more than one use with respect to the same predicate, haplogy of SIE in verb chains and finally structure sharing of SIE under coordination. The proposed analysis assumes that SIE 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 SIE.

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.