Lexical Analysis of Polish Multifunctional Reflexive Marker

Anna Kupść
Polish Academy of Sciences and Université Paris 7

1 Introduction

The aim of this paper is to argue for and propose a lexical analysis of (all types of) the Polish reflexive marker (RM) *się* 'self'.[1]

As in many Slavic and Romance languages, Polish RM is multifunctional. The anaphoric RM (ARM), unlike other types of RM, is usually treated as a counterpart of the anaphoric complement *siebie*. We provide evidence that ARM is not fully equivalent to *siebie*, either, and we propose a lexical analysis of all verbs which occur with RM. Our analysis is couched within the theoretical framework of HPSG (Head-driven Phrase Structure Grammar, cf. Pollard and Sag (1994)). We use lexical rules and lexical specification to account for various constructions with RM as well as these verb forms which obligatorily occur with RM, e.g., inherent reflexive verbs. We also show how this analysis interacts with an HPSG-encoded binding theory.

2 Properties of RM

2.1 Multifunctional RM

In its most prominent function, RM *się* is often considered a weak counterpart of the anaphoric (both reflexive and reciprocal) complement *siebie* 'oneself', cf. (1).

(1) a. reflexive
Janek$_i$ zobaczył się$_i$ / siebie$_i$ w lustrze.
John saw self$_{refl,acc}$ in mirror
'John saw himself in the mirror.'

b. reciprocal

[1] We wish to thank Anne Abeillé, Małgorzata Marciniak, Adam Przepiórkowski, Marek Świdziński and editors of this volume for very useful comments and suggestions. We benefited also from the comments of audiences at the 1999 Spring Linguistic Colloquim at the University of North Carolina and the 8th FASL Workshop. Special thanks to Tracy Holloway King for typesetting the paper in Word. The research presented in this paper was partly supported by the CRIT-2 grant.

Dzieci$_i$ *się*$_i$ / *siebie*$_i$ *straszą* *opowieściami*
children self$_{recp,acc}$ frighten stories
o *duchach.*
about ghosts
'Children frighten each other with stories about ghosts.'

Polish anaphoric pronouns, e.g., ARM and *siebie*, have only the case value assigned but are unspecified for number, gender and person. Unlike in French or English, there is no morphological agreement between an anaphoric pronoun and its nominal antecedent.

In addition to (1), Polish RM has other functions. It can be an impersonal, (2a), a middle (or inchoative), (2b), or an inherent RM, (2c).[2]

(2) a. impersonal
Czyta *się* / **siebie* *tę* *książkę* *z*
reads self$_{imp}$ this book with
przyjemnością.
pleasure
'One reads this book with pleasure.'

b. middle
Zupa *się* / **siebie* *szybko* *gotuje.*
soup self$_{midd}$ quickly cooks
'Soup cooks quickly.'

c. inherent
Jan *boi* *się* / **siebie* *Piotra.*
John fears self$_{inh}$ Peter
'John is afraid of Peter.'

In none of these functions RM can be replaced by the anaphoric complement *siebie*, unlike in (1). Verbs which occur in (2) are usually stored with *się* in the lexicon, unlike transitive verbs in (1). In subsequent sections we show, however, that the correspondence between *siebie* and ARM is not always maintained and we argue that the presence of ARM is lexically conditioned as well.

[2]See e.g., Wilczewska (1966), Saloni (1976), Kański (1984), Kubiński (1987) for various classifications of types of *się*.

2.2 RM as a Weak Form

Cardinaletti and Starke (1994) provide an apparently universal classification of pronominals. Although the overall classification is not unproblematic, some properties which distinguish ARM from the complement *siebie* may follow from the fact that ARM is a weak pronoun whereas *siebie* is a strong form.

Weak forms, in contrast to strong forms cannot be used in isolation:

(3) *Kogo widzisz na tym zdjęciu? Siebie. / *Się.*
who see-you on this picture self$_{strong}$ self$_{weak}$
'Whom do you see in this picture? Myself.'

A weak pronoun, unlike a strong one, cannot occur as a conjunct in a coordinated structure:

(4) *Jan umył siebie / *się i dzieci.*
John washed self$_{strong}$ self$_{weak}$ and children
'John washed himself and the children.'

Weak forms can never be modified:

(5) *Jan całego siebie / *się umył.*
John whole self$_{strong}$ self$_{weak}$ washed
'John washed all of himself.'

Note that in (3)–(5) the presence of RM is not excluded in general; (1a) and (6) are fully grammatical.

(6) *Jan się umył.*
John self washed
'John washed himself.'

Positions of weak forms are more restricted than those of strong ones. If postverbal, RM must be adjacent to a verb, unlike *siebie*, cf. (7) (a postverbal RM can be separated from the verb only by another weak form).

(7) *Jan umył wczoraj siebie / *się.*
John washed yesterday self$_{strong}$ self$_{weak}$
'John washed (himself) yesterday.'

RM cannot be used as a complement of a preposition, cf. (8).[3]

(8) *Położył kurtkę obok siebie / *się.*
put-he jacket next-to self$_{strong}$ self$_{weak}$
'He put the jacket next to himself.'

Note that the ungrammaticality of (8) does not follow from the fact that prepositions can never host RM. If a preposition has no complement, e.g., *dookoła* 'around' in (9), it may host RM.

(9) *Szybko dookoła się rozejrzała.*
quickly around self$_{inh}$ looked-she
'She quickly looked around.'

Also the sentence initial position of RM is excluded, cf. (10a).[4] This restriction, however, seems to follow from prosodic rather than syntactic constraints since (10b) is fully grammatical.

(10) a. *Siebie / *się wczoraj nie umył.*
self$_{strong}$ self$_{weak}$ yesterday not washed-he
'He didn't wash himself yesterday.'

b. *Piotr powiedział, że siebie / się wczoraj nie umył.*
Peter said COMP self$_{strong}$ self$_{weak}$ yesterday not washed-he
'Peter said that he hadn't washed (himself) the other day.'

According to Cardinaletti and Starke (1994), strong forms, unlike deficient (weak) forms, may have only human reference. This distinction, however, is not respected, e.g., by Polish *siebie* and ARM. In (11), both pronouns may refer to dogs.

(11) *Gdy psy$_i$ zobaczyły się$_i$ / siebie$_i$ na wzajem, zaczęły szczekać.*
when dogs saw self$_{weak}$ self$_{strong}$ each other started-they bark$_{inf}$
'When dogs saw each other in the street, they started to bark.'

[3] The only counterexample in the contemporary Polish seems to be a fixed expression *samo przez się* 'all by itself'. In Old Polish, however, weak forms after prepositions were fully correct, e.g., Klemensiewicz *et al.* (1955).

[4] In informal speech, sentence-initial positions of RM are more acceptable.

Since ARM and *siebie* successfully pass most of the tests to distinguish weak from strong forms proposed in Cardinaletti and Starke (1994), we assume here that the presented contrasts follow from this distinction.

2.3 Restrictions on Anaphoric RM

In this section we show that various verbs may have only the strong form as a complement while ARM is excluded.

As (12) illustrates, certain transitive verbs may occur only with the anaphoric complement *siebie* whereas ARM is blocked.[5]

(12) a. *Jan siebie / ?*się lubi / rozumie / kocha.*
John oneself$_{refl/recp}$ self$_{?*refl/*recp}$ likes understands loves
'John likes/understands/loves himself.'

b. *Oni siebie / się lubią / rozumieją / kochają.*
they oneself$_{refl/recp}$ self$_{??refl/recp}$ like understand love
'They like/understand/love each other/themselves.'

Such contrasts are easily explained if verbs are lexically (or semantically) specified whether they may occur with RM. The reflexive interpretation of RM is highly degraded in both (12) but a reciprocal interpretation is possible. Since the reciprocal pronoun requires a collective (plural) antecedent, it is available only in (12b).

In Polish, the presence of RM (in inherently reflexive verbs) need not exclude another nominal argument, e.g., the verb *bać się* 'be afraid' in (2c) has a genitive argument as well. If used with such verbs, the anaphoric complement cannot be substituted by ARM, cf. (13).

(13) *Jan się siebie / *się boi.*
John self$_{inh}$ oneself$_{refl,gen}$ self$_{refl,gen}$ fears
'John is afraid of himself.'

Impersonal constructions with *się*, e.g., (2a), are very productive in Polish and can be formed from various predicates, cf.

[5]This fact has been brought to our attention by Adam Przepiórkowski.

e.g., Dziwirek (1994) for an overview. If the impersonal construction is formed from a verb which may have the anaphoric complement, the strong form *siebie* is possible but ARM cannot be used, cf. (14).

(14) *Janowi łatwo się siebie / *się myje w ciepłej*
John$_{dat}$ easily self$_{imp}$ oneself$_{refl}$ self$_{refl}$ washes in warm
wodzie.
water

'John finds it easy to wash himself in warm water.'

Note that the ungrammaticality of (13) and (14) if ARM is used cannot be due to the fact that the presence of RM blocks occurrence of another anaphoric argument. The sentences with *siebie* are grammatical. Also if only one RM is present, the sentences are fully acceptable:

(15) a. *Jan się boi.*
John self$_{inh}$ fears
'John is afraid.'

b. *Janowi łatwo się myje w ciepłej wodzie.*
John$_{dat}$ easy self$_{refl/imp}$ washes in warm water
'John finds it easy to wash (himself) in warm water.'

The contrasts in (13)–(14) and the grammaticality of (15) indicate that a verb can license at most one RM whatever its function is.[6]

Polish RM exhibits so-called 'haplology'. Haplology presents an apparently paradoxical situation: one occurrence of a morpheme fulfills several functions of this morpheme. As illustrated in (16a), a single *się* serves both as an impersonal and an inherent RM (*przechadzać się* 'stroll' is an inherently reflexive verb). Several RMs cannot be used, (16b).[7]

(16) a. *Przyjemnie się nam przechadza po parku.*
nicely self$_{inh/imp}$ us$_{dat}$ strolls in park
'We enjoy strolling in the park.'

b. * *Przyjemnie się się nam przechadza po parku.*

[6] A similar observation is also made, e.g., in Rozwadowska (1992) or Rappaport (1997).

[7] For the discussion of haplology of Polish RM in other environments see, e.g., Fowler (1993), Rappaport (1997) and Kupść (1999).

As observed above, Polish verbs may license at most one RM. Therefore, one might attribute the obligatory single occurrence of RM in (13)–(15) to haplology. In contrast to (16a), however, sentences with a single RM in (15), do not have a haplology interpretation. (15a) has an 'object-drop' meaning, i.e., 'John is afraid', and does not have the intended haplology interpretation 'John is afraid of himself'. The sentence in (15b) is ambiguous. It can mean either 'John finds it easy to wash himself' (a haplology interpretation) or 'John finds it easy to wash something', washing of an unspecified object (an 'object-drop' meaning). Therefore haplology does not explain the ungrammaticality of (13)–(14) if several RMs co-occur.

Cross-linguistically, haplology is often morphologically conditioned and explained by the fact that affixes (inflectional or phrasal) cannot be doubled, e.g., Stemberger (1981), Zwicky (1987), Miller (1992). Nevertheless, a morphological explanation of haplology is not possible in the case of Polish RM. As extensively argued in Kupść (1999), Polish RM is a syntactic item (a postlexical clitic in the terminology of Miller (1992)). Unlike affixes, RM does not form a prosodic lexical unit with the host, does not have a fixed position and can be elided on a par with syntactic complements.

In §4, we provide a syntactic analysis which correctly accounts for data discussed so far, including haplology.

3 RM and Binding

In Polish, as in many other Slavic languages, binding of anaphoric pronouns is subject-oriented, see (17) (=(1)).[8]

(17) a. *Janek$_i$ zobaczył się$_i$ / siebie$_i$ w lustrze.*
John saw self$_{refl}$ in mirror
'John saw himself in the mirror.'

b. *Dzieci$_i$ się$_i$ / siebie$_i$ straszą opowieściami o duchach.*
children self$_{recp}$ frighten stories about ghosts
'Children frighten each other with stories about ghosts.'

[8]Unlike reflexive pronouns, reciprocal pronouns may be bound by a non-subject NP as well, e.g., Marciniak (1999). We ignore this fact here.

If RM has other functions, see (2), it seems to be only a syntactic marker rather than an anaphoric pronoun. Although it is not clear to us whether these RMs should be subject to binding principles, anaphoric co-arguments of RM in inherent reflexives and middles are bound by the subject, cf. (18a) and (18b), respectively.

(18) a. *Jan$_i$ bał się siebie$_i$.*
John self$_{inh}$ feared oneself
'John was afraid of himself.'

b. *Gałąź$_i$ złamała się pod ciężarem swoich$_i$ owoców.*
branch broke self$_{midd}$ under weight self's fruits
'The branch broke under the weight of its fruits.'

In impersonal constructions, cf. (19), binding of anaphoric complements is less straightforward.

(19) a. *Trudno rozmawia się z innymi$_{*i}$ o sobie$_i$.*
difficult talks self$_{imp}$ with others about self
'It is difficult to talk to other people about oneself.'

b. *Trudno się mi$_i$ rozmawia z innymi$_{*i}$ o sobie$_i$.*
difficult self$_{imp}$ me$_{dat}$ talks with others about self
'It is difficult for me to talk to other people about myself.'

Unlike in (17) or (18), the anaphoric complement in (19a) refers to some unspecified (human or animate) object. In (19b), the binder is overtly expressed but it is a dative rather than a nominative phrase.

The specification of the syntactic subject (or a binder) in impersonal constructions has been a long-standing issue of Polish grammar, cf. e.g., Wilczewska (1966), Saloni (1976), Kański (1984), Kubiński (1987). Saloni (1976) argues that in such constructions there is a nominative subject, namely *się*. Hence in (19a), RM can be an overt binder for the anaphoric *sobie*. In (19b), however, the complement is coindexed with the dative phrase *mi* rather than with RM. Note that the main difference between both sentences is the presence of NP[dat] in (19b) and lack thereof in (19a). Due to these similarities, RM in both constructions should have the same status. If RM is the syntactic subject in both constructions, then the obligatory coindexation with NP[dat] in (19b) could be obtained only if RM and NP[dat] were coindexed. Observe that if RM is the subject, it (locally) commands NP[dat]. Thus (pronominal) NP[dat] in (19b) would be (locally) bound ((locally) commanded by and coindexed with *się*), in violation of Principle B (or Principle C if NP[dat] is an R-expression).

The account we propose is quite different. We assume that impersonal constructions are subjectless, i.e., there is no overt subject in any of (19). Neither RM nor the dative NP in (19) are syntactic (surface) subjects. We assume that in impersonal constructions the binder need not be overtly realised but it is always present at a different (semantic) level. Once realised, however, it surfaces as NP[dat].

A similar distinction between covert and overt subjects has been used in Marciniak (1999) to account for binding of arguments of participial impersonals:

(20) *Rozmawiano o Tomku*$_{*i}$ *ze sobą*$_i$.
talked$_{particip_imp}$ about Tom with self
'They were talking to each other about Tom.'

In (20), no overt subject can be present but the anaphoric pronoun must be bound by an implicit argument (a syntactically unrealised subject). Coindexation with any other argument is ungrammatical.

4 HPSG Account

The analysis presented in this section is mostly based on that of Kupść (1999), extended to account for binding facts discussed in §3.

4.1 General Assumptions

As stated in §2.3, Polish RM is a syntactic item. A schematic lexical entry for *się* is given in (21).

(21)
$$\begin{bmatrix} \textit{word} & \\ \text{PHON} & \langle \text{się} \rangle \\ \text{SS} & \begin{bmatrix} \textit{clitic} & \\ \text{LOC} & \begin{bmatrix} \textit{local} & \\ \text{CAT}|\text{HEAD} & \textit{noun} \\ \text{CONT} & \begin{bmatrix} \textit{ana} & \\ \text{INDEX} & \textit{index} \\ \text{RESTRICTION} & \textit{set\_psoa} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

This lexical entry comprises all types of RMs distinguished in §2.1. We assume that there is only one lexical entry for *się*, e.g., contrary to Saloni (1976). Various functions of RM, exemplified in (1) and (2), will be established by (semantics of) a verb. Referential properties of nominal objects (hence RM as well) are encoded as the value of the INDEX attribute, i.e., *index*, which can be split into *referential* and *non-referential* subsorts. The RESTRICTION (RESTR) attribute allows one to restrict content of nominals, e.g., require NPs to denote only animate objects.

We follow Sag (1997), Miller and Sag (1997), Abeillé *et al.* (1998) and split *synsem* objects into subsorts (*synsem*s represent syntactico-semantic properties of linguistic objects). In order to distinguish RM from other NPs, we introduce subsorts *clitic* and *non-clitic*, correspondingly.[9]

We assume that RM is subcategorized for by a verb on a par with other complements.[10] In HPSG, the subcategorization frame is encoded via VALENCE (VAL) attributes. A schematic lexical entry for the verb *boi się* 'fears' is given in (22).

(22)
$$\begin{bmatrix} \textit{word} & \\ \text{PHON} & \langle \text{boi} \rangle \\ \text{SS}|\text{LOC}|\text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \boxed{1} \langle \text{NP}[\textit{str}] \rangle \\ \text{COMPS} & \boxed{2} \langle \textit{ana-cl}, \text{NP}[\textit{lgen}] \rangle \end{bmatrix} \\ \text{ARG-ST} & \boxed{1} \oplus \boxed{2} \end{bmatrix} \end{bmatrix}$$

The VAL attributes contain information about arguments of the (lexical or phrasal) head. Values of these attributes are lists of *synsem*s. The ARG-ST (ARGUMENT-STRUCTURE) attribute is specified only for *word*s and is the locus of HPSG binding theory, cf. §4.6. ARG-ST and VAL are related. In most cases, ARG-ST

[9] The Polish *się* is a syntactic clitic, cf. Kupść (1999). Thus, *clitic* is a subsort of the *canonical synsem* (on a par with *non-clitic*), unlike French morphological clitics in Miller and Sag (1997).

[10] Unlike Kupść (1999), we do not use a separate attribute for RM.

is just concatenation of VAL|SUBJ and VAL|COMPS lists (in this order) but this relationship need not always be preserved, cf. Manning and Sag (1999). In a more elaborated account, once the precise relationship between ARG-ST and VAL is established, it is sufficient to specify only the ARG-ST value in the lexicon as VAL can be deduced. Since we do not formalise the principle which relates them, we explicitly specify both VAL and ARG-ST in the paper.

We use *ana-cl* as the abbreviation for the SYNSEM (SS) value given in the lexical entry for *się* in (21). Hence, RM is a (subcategorized for) nominal object of the sort *clitic* with (referential or not) anaphoric properties ([CONT *ana*]).

We assume that RM is present on ARG-ST, cf. (22). Therefore if RM bears a referential index, it will obey binding principles, cf. §4.6. Also if case assignment is encoded on ARG-ST, as proposed in Przepiórkowski (1999), the (structural) case value of RM will be resolved by the case principle.[11]

As discussed in §2.3, a verb can license at most one RM. The following constraint ensures that ARG-ST contains at most one *ana-cl*:[12]

(23) $$\begin{bmatrix} \textit{word} \\ \text{SS|LOC|CAT|ARG-ST} \quad \boxed{1}\textit{ne\_list}(\textit{ana-cl}) \bigcirc \textit{list}(\textit{non-ana-cl}) \end{bmatrix} \rightarrow \boxed{1}\langle \textit{ana-cl} \rangle$$

This constraint is implicational and does not force verbs to occur with *się*. It says that **if** *ana-cl* is present on the ARG-ST list, this is the only *ana-cl* element of this list. We use *non-ana-cl* as the abbreviation for a *synsem* object distinct from *ana-cl*.

Since RM is a syntactic item, traditional HPSG syntactic principles, i.e., the Valence Principle and Immediate Dominance Schemata, ensure its realisation. No additional mechanisms are necessary.

Having introduced these general assumptions of our analysis, we may proceed to the account of data discussed in previous sections.

[11] We follow Przepiórkowski (1996) and adopt lexical vs. structural case dichotomy for Polish.

[12] The '$\bigcirc$' indicates the 'shuffle' relation introduced in Reape (1992). This relation operates on lists and 'shuffles' their elements in such a way that the relative order of elements of original lists is preserved.

4.2 Inherent Reflexives

Inherent reflexive verbs do not have a non-reflexive counterpart and RM is specified directly in the lexical entry. The lexical entry for *boi się* 'fears' given in (22) is repeated as (24).

(24)
$$
\left[\begin{array}{ll}
\textit{word} & \\
\text{PHON} & \langle \text{boi} \rangle \\
\text{SS}|\text{LOC}|\text{CAT} & \left[\begin{array}{ll} \text{HEAD} & \textit{verb} \\ \text{VAL} & \left[\begin{array}{ll} \text{SUBJ} & \boxed{1}\langle \text{NP}[\textit{str}]\rangle \\ \text{COMPS} & \boxed{2}\langle \textit{ana-cl}, \text{NP}[\textit{lgen}]\rangle \end{array}\right] \\ \text{ARG-ST} & \boxed{1}\oplus\boxed{2} \end{array}\right]
\end{array}\right]
$$

Due to the constraint in (23) and the lexical entry in (24), multiple RMs in (13) are correctly ruled out. The verb has two elements on the COMPS list: *się* and a genitive NP. The realisation of the nominal genitive argument as another *się* is correctly blocked by (23). On the other hand, nothing prevents the co-occurrence of several anaphoric arguments, i.e., (13) is licensed as desired. There is no principle that forces haplology in general. Hence, an 'object-drop' reading in (15a) is predicted: only RM is realised while the complement is missing.

4.3 Verbs with ARM

Verbs such as *zobaczył* 'saw-he' or *straszą* 'frighten-they' in (1) can occur both with ARM or *siebie*. We assume that their COMPS value is left unspecified, cf. (25), i.e., the complement NP can be of any *synsem* sort. Hence it can be realised either as *ana-cl* or any other accusative NP. This correctly predicts that in (1) both forms, *się* and *siebie*, are possible.

(25)
$$
\left[\begin{array}{ll}
\textit{word} & \\
\text{PHON} & \langle \text{zobaczył} \rangle \\
\text{SS}|\text{LOC}|\text{CAT} & \left[\begin{array}{ll} \text{HEAD} & \text{verb} \\ \text{VAL} & \left[\begin{array}{ll} \text{SUBJ} & \boxed{1}\langle \text{NP}[\textit{str}]\rangle \\ \text{COMPS} & \boxed{2}\langle \text{NP}[\textit{str}]\rangle \end{array}\right] \\ \text{ARG-ST} & \boxed{1}\oplus\boxed{2} \end{array}\right]
\end{array}\right]
$$

For verbs such as in (12), we stipulate here that their only NP complement cannot be the *reflexive* clitic, i.e., *reciprocal* or *non-clitic*:*refl*. This rules out ungrammatical forms with the reflexive RM. Certainly, a more principled (semantic) explanation should be proposed since similar restrictions hold in other languages, e.g., Serbian/Croatian, according to Wayles Browne (p.c.). We leave this issue for future study.

4.4 Impersonal Verbs

We assume that impersonal verbs with *się* are obtained via a lexical rule[13] given schematically in (26).[14]

(26)
$$
\left[\begin{array}{ll}
\textit{word} & \\
\text{SS}|\text{LOC}|\text{CAT} & \left[\begin{array}{ll} \textit{category} & \\ \text{HEAD} & \textit{verb} \\ \text{VAL} & \left[\begin{array}{ll} \textit{val} & \\ \text{SUBJ} & \boxed{6}\langle \text{NP}[\textit{str}]{:}\boxed{1}\rangle \\ \text{COMPS} & \boxed{4}\text{list}(\textit{ana-cl})\oplus\boxed{2}\text{list}(\textit{non-ana-cl}) \end{array}\right] \\ \text{ARG-ST} & \boxed{6}\oplus\boxed{4}\oplus\boxed{2} \end{array}\right]
\end{array}\right]
$$
$$
\mapsto \left[\begin{array}{ll}
\text{SS}|\text{LOC}|\text{CAT} & \left[\begin{array}{ll} \text{HEAD}|\text{VFORM} & \textit{fin}_{3rd\_neut} \\ \text{VAL} & \left[\begin{array}{ll} \text{SUBJ} & \langle\rangle \\ \text{COMPS} & (\boxed{3}\langle \text{NP}[\textit{ldat}]{:}\boxed{1}\rangle\oplus)\boxed{5}\langle \textit{ana-cl}\rangle\oplus\boxed{2} \end{array}\right] \\ \text{ARG-ST} & \boxed{3}\oplus\boxed{5}\oplus\boxed{2} \end{array}\right]
\end{array}\right]
$$

Although impersonal constructions are quite productive in Polish, they are not allowed with all verbs. First, impersonals are possible only with verbs which require a human (or animate) subject. This can be obtained by restricting the value of input's VAL|SUBJ to animate NPs (via the CONT|RESTR value). Second, certain psych-verbs, e.g., *bać się* 'to be afraid' do not have impersonal forms with *się* although they do require an animate subject. Thus (26) has to be further restricted. Precise semantic constraints are not clear to us at the moment and we concentrate on syntactic issues mainly.

(26) provides only this information which distinguishes input and output verbs. The ARG-ST lists of the input and output verbs can be distinct if the input verb is not reflexive, i.e., if there is no *ana-cl* on the ARG-ST list. Note that the COMPS value of the input contains a list of *ana-cl*, i.e., $\boxed{4}$ may be either empty or non-empty. Distinct VAL values of input and output verbs correctly block recursive application of the rule. The output verb must occur in third neuter singular form. This verb form is predicted by the subject-predicate agreement principle proposed in Czuba and Przepiórkowski (1995).

The lexical rule in (26) licenses only one RM in the output verb. Irrespectively of the input specification of the verb only one RM results. If $\boxed{4}$ is not empty, (26) correctly rules out

[13] See Monachesi (1995) uses lexical rules to account for the Italian reflexive marker *si*.

[14] NP:$\boxed{1}$ abbreviates an NP with the [SS|LOC|CONTENT $\boxed{1}$] path value.

(14) and (16b) and predicts that only one RM appears in (15b) and (16a). The optionality of the dative argument in impersonal constructions, cf. (2a) vs. (18a) and (16a), is indicated by the brackets around $\boxed{3}$. Note, however, that NP[dat] is always present on the ARG-ST. This fact will be crucial for binding in §4.6.

Our analysis accounts also for the ambiguity of (15b). We assume that *myje* 'washes' has the lexical entry analogous to (25), i.e., the complement can be either realised as *ana-cl* or any other NP. If in the input of (26) a non-reflexive verb is used, an 'object-drop' reading results. Only the impersonal RM is realised while the complement of the transitive *myje* 'washes' remains unrealised, i.e., $\boxed{2}$ is non-empty. However, if the input verb occurs with ARM, haplology results. The input's NP complement is realised as *ana-cl* and $\boxed{2}$ is the empty list. Hence in the output of (26), RM in $\boxed{5}$ serves both as an impersonal RM and ARM.

4.5 Middles

We propose a (schematic) lexical rule in (27) to relate transitive verbs and their 'middle' counterparts.

(27)
$$\begin{bmatrix} \textit{word} & \\ \textsc{ss}|\textsc{loc}|\textsc{cat} & \begin{bmatrix} \textit{category} & \\ \textsc{head} & \textit{verb} \\ \textsc{val} & \begin{bmatrix} \textit{val} & \\ \textsc{subj} & \boxed{2}\langle \text{NP}[\textit{str}]\rangle \\ \textsc{comps} & \boxed{3}\langle \text{NP}[\textit{str}]{:}\boxed{1}\rangle \end{bmatrix} \\ \textsc{arg-st} & \boxed{2}\oplus\boxed{3} \end{bmatrix} \end{bmatrix}$$
$$\mapsto \begin{bmatrix} \textsc{ss}|\textsc{loc}|\textsc{cat} & \begin{bmatrix} \textsc{val} & \begin{bmatrix} \textsc{subj} & \boxed{4}\langle\text{NP}[\textit{str}]{:}\boxed{1}\rangle \\ \textsc{comps} & \boxed{5}\langle\textit{ana-cl}\rangle\end{bmatrix} \\ \textsc{arg-st} & \boxed{4}\oplus\boxed{5}\end{bmatrix}\end{bmatrix}$$

As in the case of impersonals, there are semantic restrictions on formation of middle constructions. For example, (27) should mainly apply to verbs which have a non-animate complement. Thus, the CONT value of input's complement should be restricted. Middle constructions in Polish are not very productive and much stronger semantic restrictions should apply. At present, we assume that middle forms are derived by the lexical rule (rather than enumerated in the lexicon) and we discuss briefly syntactic aspects of (27).

The effect of (27) is to change the complement of the (transitive) input verb into the output's subject and add RM. We do not impose any restrictions on the COMPS value of the input verb in spite of the fact that middles cannot be formed from verbs which occur with RM. The derivation of middle forms from such verbs, however, will be blocked. In Polish, anaphoric NPs are not subjects. The SUBJ value of the output is non-empty and NP's CONTENT value, i.e., $\boxed{1}$, is token-identical with that of the complement of the input. Thus, the output's subject has the same semantic properties as the complement of the input. If the latter is anaphoric, the general principle of Polish blocks its realisation as the subject. This also blocks recursive application of (27) (the output's VAL|COMPS is *ana-cl*).

4.6 Binding Theory

As mentioned in §4.1, HPSG binding theory is encoded on the ARG-ST list, cf. Abeillé *et al.* (1998), Manning and Sag (1999). In HPSG, binding theory is defined in terms of the relative obliqueness of grammatical functions rather than in purely configurational terms. Pollard and Sag (1994) define o-command (obliqueness-command), (28), and o-binding (obliqueness-binding), (29), relations which serve for the definition of binding principles, (30).[15]

(28) Let X and Y be *synsem* objects, X referential. Then X *locally o-commands* Y just in case X is less oblique than Y (X precedes Y on the ARG-ST list).

(29) X *locally o-binds* Y just in case X and Y are coindexed and X locally o-commands Y.

(30) PRINCIPLE A

A locally o-commanded anaphor must be locally o-bound.

Such a formulation of anaphoric binding accounts for these cases when 1) there is a local binder, i.e., a preceding element on ARG-ST exists, and 2) the local binder is referential. Otherwise, anaphoric elements present on ARG-ST are exempt from (30) but can be bound on other non-syntactic grounds.

[15] We present only these definitions of Pollard and Sag (1994) which are relevant for our present discussion.

In Polish, (reflexive) anaphora are subject-oriented[16] and the definition of the binder must be more specific. We quote below relevant definitions of the binding theory for Polish presented in Marciniak (1999).

(31) Let X and Y be *synsem* objects, where X is referential. Then X *locally s-commands* Y if X and Y are elements of the same ARG-ST list and X is the first element of this ARG-ST list; ...

(32) X *locally s-binds* Y just in case X and Y are coindexed and X locally s-commands Y.

(33) PRINCIPLE A

A reflexive anaphor must be locally s-bound.

As mentioned in §4.1, the ARG-ST value is usually the concatenation of values of VAL attributes in the canonical oblique-based order, i.e., the subject (VAL|SUBJ) plus complements (VAL|COMPS), e.g., (24) and (25). For example, the ARG-ST list for *boi się* 'fears' in (18a) is given as follows:

(34) ARG-ST $\langle$NP$_i$[*str*]:*npro*, *ana-cl*, NP$_i$[*lgen*]:*ana*$\rangle$

The definitions in (31)–(33) correctly predict that *siebie* refers to *Jan* 'John'. The (referential) complement *siebie*, indicated as NP$_i$[*lgen*]:*ana*, and the subject *Jan*, i.e., NP$_i$[*str*]:*npro*, are elements of the same ARG-ST list. According to (31), the latter locally s-commands the former. Since these NPs are coindexed, *Jan* locally s-binds *siebie*, cf. (32), and the Principle A in (33) is satisfied as desired. The binding of anaphora in (17) (and (18b)) is obtained analogously.[17]

This theory accounts also for binding in impersonal constructions. The ARG-ST value for *rozmawia* 'talks' in (19) is given in (35).

(35) ARG-ST $\langle$NP$_i$[*ldat*]:*npro*, *ana-cl*, PP$_{*i}$[z]:*npro*, PP$_i$[o]:*ana* $\rangle$

The dative phrase is the **first** element on ARG-ST. As mentioned above, NP[dat] need not be overtly realised. For example, in (2a) and (19a), there would be no corresponding NP[dat] on the

[16] See fn. 8.

[17] The anaphoric pronoun in (18b) is not a direct argument of the verb. If we apply remaining (not quoted here) clauses of the *local s-command* definition in Marciniak (1999), binding is correctly obtained.

VAL|COMPS lists. Since NP[dat] is always present on ARG-ST of impersonal verbs, the anaphoric complements, e.g., PP *o sobie* 'about oneself' in (19) are locally s-bound. In (19a), NP[dat] denotes an unspecified (animate) object while in (19b) this object is overtly expressed as *mi*. Due to definitions in Marciniak (1999), NP[dat] locally s-commands *sobie*. Since these NPs are coindexed, NP[dat] locally s-binds the anaphora, (32), and (33) is satisfied as desired.

Note that if RM were the first element of ARG-ST in (35), the correct binding in (19b) would not be predicted. If RM was referential, then *się* (rather than NP[dat]) would be the local s-binder for the anaphora. One can suggest that NP[dat] might become a local s-binder indirectly, i.e., if it were coindexed with (referential) RM. Note, however, that in this case (non-anaphoric) NP[dat] becomes (locally) s-bound, contrary to Principle B and C. On the other hand, if the first element of ARG-ST was a non-referential RM, binding of PP *o sobie* with *mi* would be exempt from our binding theory. Therefore, s-binding with NP[dat] might be predicted on different grounds. If (33) does not hold, however, nothing prevents illegal coindexation of the anaphora with PP *z innymi*.

5 Conclusions

In this paper we presented a lexical analysis of the Polish reflexive marker *się*. Such an analysis allows us to account for these cases when the correspondence between the anaphoric complement *siebie* and ARM is not maintained. Our approach accounts also for so-called haplology of RM. Although Polish verbs may license at most one RM, haplology does not always result. Our analysis correctly predicts this fact since haplology is accounted for by a lexical rule which derives impersonals. Due to the lexical specification of the ARG-ST value for impersonals, we capture problematic bindings in impersonal constructions.

Although the presented analysis correctly deals with the data, it is mostly syntactic. For example, we have to stipulate the subcategorization frame for verbs which do not allow ARM whereas a more principled semantic explanation should be proposed. Also lexical rules we sketched reflect mostly syntactic properties of discussed constructions. Semantic restrictions have not been incorporated. We hope to resolve these issues in future study.

References

Abeillé, A., Godard, D., and Sag, I. A. (1998). Two kinds of composition in French complex predicates. In E. Hinrichs, A. Kathol, and T. Nakazawa, editors, *Complex Predicates in Nonderivational Syntax*, volume 30 of *Syntax and Semantics*, pages 1–41. Academic Press, New York.

Borsley, R. D. and Przepiórkowski, A., editors (1999). *Slavic in Head-Driven Phrase Structure Grammar*. CSLI Publications, Stanford.

Cardinaletti, A. and Starke, M. (1994). The typology of structural deficiency on the three grammatical classes. *University of Venice Working Papers in Linguistics*, **4**(2), 41–109.

Czuba, K. and Przepiórkowski, A. (1995). Agreement and case assignment in Polish: An attempt at a unified account. Technical Report 783, Institute of Computer Science, Polish Academy of Sciences.

Dziwirek, K. (1994). *Polish Subjects*. Garland, New York & London.

Fowler, G. (1993). A syntactic account of derivational *-sja* in Russian. In R. Maguire and A. Timberlake, editors, *American Contributions to the Eleventh International Congress of Slavists*, pages 270–284. Slavica Publishers, Columbus, Ohio.

Kański, Z. (1984). *Arbitrary Reference and Reflexivity: A Generative Study of the Polish Pronoun* się *and its English Equivalents*. Ph.D. thesis, University of Silesia, Katowice.

Klemensiewicz, Z., Lehr-Spławiński, T., and Urbańczyk, S. (1955). *Gramatyka Historyczna Języka Polskiego*. Państwowe Wydawnictwo Naukowe.

Kubiński, W. (1987). *Reflexivization in English and Polish: An Arc Pair Grammar Analysis*, volume 178 of *Linguistische Arbeiten*. Niemeyer, Tübingen.

Kupść, A. (1999). Haplology of the Polish reflexive marker. In Borsley and Przepiórkowski (1999), pages 91–124.

Manning, C. D. and Sag, I. A. (1999). Dissociations between argument structure and grammatical relations. In Webelhuth *et al.* (1999), pages 63–78.

Marciniak, M. (1999). Toward a binding theory for Polish. In Borsley and Przepiórkowski (1999), pages 125–147.

Miller, P. H. (1992). *Clitics and Constituents in Phrase Structure Grammar*. Garland, New York.

Miller, P. H. and Sag, I. A. (1997). French clitic movement without clitics or movement. *Natural Language and Linguistic Theory*, **15**, 573–639.

Monachesi, P. (1995). *A Grammar of Italian Clitics*. Ph.D. thesis, Tilburg University, Tilburg. ITK Dissertation Series 1995-3 and TILDIL Dissertation Series 1995-3.

Pollard, C. and Sag, I. A. (1994). *Head-driven Phrase Structure Grammar*. Chicago University Press, Chicago.

Przepiórkowski, A. (1996). Case assignment in Polish: Towards an HPSG analysis. In C. Grover and E. Vallduví, editors, *Studies in HPSG*, volume 12 of *Edinburgh Working Papers in Cognitive Science*, pages 191–228. Centre for Cognitive Science, University of Edinburgh.

Przepiórkowski, A. (1999). On case assignment and "adjuncts as complements". In Webelhuth *et al.* (1999), pages 231–245.

Rappaport, G. C. (1997). Clitics as features: A non-semiotic approach. To appear in: American Contributions to the International Congress of Slavists. Version of 7th March 1997.

Reape, M. (1992). *A Formal Theory of Word Order: A Case Study in West Germanic*. Ph.D. thesis, University of Edinburgh.

Rozwadowska, B. (1992). *Thematic constraints on selected constructions in Polish and English*. Wydawnictwa Uniwersytetu Wrocławskiego, Wrocław.

Sag, I. A. (1997). English relative clause constructions. *Journal of Linguistics*, **33**, 431–483.

Saloni, Z. (1976). *Cechy Składniowe Polskiego Czasownika*. Ossolineum, Wrocław.

Stemberger, J. P. (1981). Morphological haplology. *Language*, **57**, 791–817.

Webelhuth, G., Koenig, J.-P., and Kathol, A., editors (1999). *Lexical and Constructional Aspects of Linguistic Explanation*. CSLI Publications, Stanford.

Wilczewska, K. (1966). Czasowniki zwrotne we współczesnej polszczyźnie. *Towarzystwo Naukowe w Toruniu, Prace Wydziału Filologiczno-Filozoficznego*, **XVII**(3).

Zwicky, A. M. (1987). Suppressing the Zs. *Journal of Linguistics*, **23**, 133–148.