Preposition-Pronoun Contraction in Polish

Beata Trawiński
Eberhard-Karls-Universität Tübingen
Sonderforschungsbereich 441
Nauklerstraße 35
D–72074 Tübingen
trawinski@sfs.uni-tuebingen.de

Abstract

This paper provides a lexicalist formal description of preposition-pronoun contraction (PPC) in Polish, using the theoretical framework of HPSG. Considering the behavior of PPC with respect to the prosodic, categorial, syntactic and semantic properties, the assumption can be made that each PPC is a morphological unit with prepositional status. The crucial difference between a PPC and a typical preposition consists, besides the phonological form, in the valence properties. While a typical preposition realizes its complement externally via general constraints on phrase structure, the realization of a PPC argument is effected internally by virtue of its lexical entry. Here, we will provide the appropriate implicational lexical constraints that license both typical Ps and PPCs.

1 Introduction

Polish prepositions exhibit the ability to coalesce with the third person pronouns they select for. In (1), two examples of a preposition-pronoun amalgamation in Polish are given. We will refer to this phenomenon as preposition-pronoun contraction (PPC).[1]

(1) a. na niego 'on him' → nań 'on_him'
b. w niego 'in him' → weń 'in_him'

In spite of the fact that PPCs are commonly used in contemporary written Polish, they are often neglected by Polish grammars, which, if they point out this phenomenon at all, do not attempt to systematize or adequately describe it.[2] Nevertheless, the treatment of PPCs relates to some interesting and non-trivial issues, in particular to the question of whether PPCs pertain exclusively to the phonological or, more precisely, prosodic domain, or whether they are also subject to any syntactic processes. Besides empirical issues, analytical questions should be taken into consideration, e.g., whether PPC should be analyzed in terms of affixation or cliticization.

This paper aspires to answer these questions and to provide a formal description of Polish PPC suitable for machine processing. We will use Head-Driven Phrase Structure Grammar in the tradition of (Pollard and Sag, 1994) as a theoretical framework to formalize empirical generalizations.

2 The Phenomenon

2.1 Case, Gender and Number Restrictions

According to traditional approaches to PPC, only genitive and accusative singular masculine third per-

[1] We use the notion contraction here as a neutral term unrelated to phonology or morphology.

[2] Cf. (Lehr-Splawiński and Kubiński, 1952), (Szober, 1969), (Świdziński, 1992), (Saloni and Świdziński, 1998), (Bąk, 2004), which make no mention of PPC in Polish at all, and (Wieczorkiewicz et al., 1957), (Doroszewski and Wieczorkiewicz, 1972), (Grzegorczykowa et al., 1984), (Nagórko, 1996), (Wróbel, 2001), which merely allude to this phenomenon.

son pronouns can contract with prepositions to form PPCs (cf. e.g., (Saloni, 1981)).[3] Below, examples of PPs are given that contain genitive and accusative singular masculine nouns (proper names and pronouns), and their corresponding PPCs.

(2) a. Maria pracuje dla Piotra / dla niego / dlań od roku.
Maria works for Piotr$_{gen}$ / for him$_{gen}$ / for_him for year
'Maria has been working for Piotr / for him for one year.'

b. Jan zapłacił za obiad / za niego / zań gotówką.
Jan payed for dinner$_{gen}$ / for it$_{gen}$ / for_it cash
'Jan payed for the dinner / for it in cash.'

The examples in (3) seem to support the assumption that only genitive and accusative singular masculine third person pronouns can contract with prepositions. Neither dative nor instrumental pronouns are possible within PPCs.[4]

(3) a. Maria pochyliła się ku dziecku / ku niemu / *kuń troskliwie.
Maria leaned RM towards child$_{dat}$ / towards it$_{dat}$ / towards_it
'Maria leaned towards the child / towards it.'

b. Anna szła przed Piotrem / przed nim / *przedeń obrażona.
Anna walked before Piotr$_{instr}$ / before him$_{instr}$ / before_him offended
'Anna walked offended before Piotr / before him.'

However, contrary to commonly accepted claims, PPCs seem to be possible with locative pronouns as well. Selected examples taken from the IPI PAN Corpus[5] (4a) and the Internet via Google (4b)[6] are given below.

(4) a. Polskie przedsiębiorstwa przemysłowe najlepiej prosperują w Mazowieckiem i Śląskiem. Pod koniec sierpnia br., uzyskało weń zatrudnienie łącznie ponad 1 mln osób [...]
Polish companies industrial best prosper in Mazowieckie and Śląskie by end August of this year gained in_them$_{loc}$ employment altogether above 1 million persons [...]
'Polish industrial companies best prosper in Mazowieckie and Śląskie. By late August of this year, a total of over 1 million people gained employment there.'

b. Taką wiedzę trzeba odpowiednio zdobyć, a to nie jest tak łatwe jak przeczytanie księgi o potworach i zdobycie ońwiedzy.
such knowledge one must appropriately gain and it not is as easy as reading book about monsters and acquiring about_them$_{loc}$ knowledge
'One must gain such a knowledge in an appropriate way and it is not as easy as to gain knowledge about monsters by reading a book about them.'

Note however that native speaker judgments differ in the use of locative pronouns within PPCs. Although we tend to assume that locative third person pronouns are also able to contract with prepositions and although we also consider the commonly assumed restrictions on the gender and number of pronouns contracting with Ps to be too strong, we will temporarily adopt the common claim that only masculine (human and non-human) singular third person pronouns can contract with Ps. Since the verification

[3] (Doroszewski and Wieczorkiewicz, 1972) even claim that PPCs are not allowed in the case of Ps requiring any other cases than accusative. To that effect, PPCs in (2a) should be considered ungrammatical. However, according to the latest dictionaries of the Polish language such as (Dubisz, 2003) or (Bańko, 2000), as well as on the basis of corpus evidence, we consider prepositions contracted with genitive third person pronouns fully grammatical.

[4] The abbreviation RM stands for a reflexive marker.

[5] See (Przepiórkowski, 2004) for details on the IPI PAN Corpus. The corpus URL: `http://www.korpus.pl`.

[6] `http://forum.inkluz.pl/post-1709.html`

of the traditional assumptions should follow detailed examination of the data, we will leave this issue for future work.

2.2 Categorial, Syntactic and Semantic Properties

As the examples in (5) illustrate, PPCs cannot combine with any further NPs. This fact indicates that they are syntactically saturated expressions corresponding to ordinary PPs.

(5) a. Maria pracuje dlań (*Piotra) od roku.
Maria works for_him Piotr$_{gen}$ for year
'Maria has been working for him for one year.'

b. Anna robi zań (*Jana) wszystko.
Anna does instead of_him Jan$_{acc}$ everything
'Anna does everything instead of him.'

While the examples in (5), where PPCs act as adjuncts, may suggest the possibility that PPCs are adverbs, the data in (6) seem to argue for a prepositional status of PPCs.[7]

(6) a. Piotr zgubił prawo jazdy, ale jakoś obywa się *(bezeń).
Piotr lost license driving but somehow dispense RM without_it
'Piotr has lost his driving license, but he can somehow do without it.'

b. Anna namawia Jana na ślub, ale wciąż nie może *(nań) skutecznie wpłynąć.
Anna persuade Jan for marriage but still not can on_him successfully affect
'Anna has been trying to persuade Jan to marry her, but she still cannot successfully move him.'

Here particular PPCs are selected by predicates which seem to always require PP complements headed by specific prepositions. As we can see in (6), PPCs can occur unconstrained instead of corresponding PPs as obligatory complements of these predicates. Omitting PPCs in the sentences above causes ungrammaticality. These observations, supported by the fact that the phonological form of each PPC contains a string identical to that of the corresponding preposition, possibly indicate that PPCs have the same syntactic and categorial status as the corresponding PPs, thus excluding the otherwise conceivable possibility of treating PPCs as a type of adverb or pronoun.[8]

The data in (7) could be considered as an apparent argument for treating PPCs as pronouns.[9]

(7) a. Maria$_i$ umówiła się z Piotrem$_j$ i czekała nań$_{*i\,/\,j}$ w mieście.
Maria made an appointment RM with Piotr and waited for_her / him in city
'Maria made an appointment with Piotr and waited for him in the city.'

b. [Anna i Maria]$_i$ zapewniły Jana$_j$, że *pro*$_i$ wierzą weń$_{*i\,/\,j}$.
Anna and Maria assured Jan that *pro* believe in_each other / him
'Anna and Maria assured Jan that they believe in him.'

c. Anna$_i$ obiecała Janowi$_j$ PRO$_i$ troszczyć się oń$_{*i\,/\,j}$.
Anna promised Jan PRO care RM for_her / him
'Anna promised Jan to care for him.'

[7] The notation *(X) as used in (6) implies that the presence of X is necessary for the grammaticality of the sentence.

[8] Cf. the traditional treatment of similar expressions in German. In German, a number of prepositions can combine with *da*, *wo* and *hier* expressions forming morphologically complex words: *dafür* 'for it', *wofür* 'for what', *hierfür* 'therefor'. These expressions are treated in traditional grammars of German as pronominal adverbs (cf. (Duden IV, 1984)), prepositional adverbs (cf. (Engel, 1996)), pronouns (cf. (Helbig and Buscha, 1993)) or P-pronouns (cf. (Fries, 1988)). For the reasons explained in (6), we suggest to handle this type of German expressions similarly to Polish PPCs, that is as single word prepositional phrases.

[9] Note, however, that in spite of the fact that the PPC *oń* 'for' as used in (7c) and (7d) are provided by dictionaries of contemporary Polish such as (Dubisz, 2003) or (Bańko, 2000), not all native speakers of Polish accept this PPC.

d. Jan$_i$ kazał Annie$_j$ PRO$_j$ troszczyć się
Jan made Anna PRO care RM
ońi / *j.
for_him / her
'Jan made Anna care for him.'

Here PPCs seem to behave like ordinary personal pronouns controlled by referential NPs. Note that the anaphoric reading is not possible in any of these sentences, nor is the reflexive (cf. the subject control construction in (7c) and the object control construction in (7d)), or the reciprocal reading (cf. (7b)).

However, the reference data in (7) refer to semantic properties of PPCs and are not the decisive factor in determining their categorial status and establishing their syntactic characteristics. They merely indicate that the semantic representation of PPCs must contain an index. This also applies to ordinary PPs.

2.3 Prosodic Properties

In Polish, neither modification nor coordination allow phonologically / prosodically deficient entities.[10] A prosodically deficient entity cannot occur sentence initially, either.

As we can see in the examples below, no modification (cf. (8a)), coordination (cf. (8b)) and occurrence in sentence initial position (cf. (8c)) is allowed in structures involving PPCs.

(8) a. Jan użył cytatu pochodzącego od
Jan used citation coming from
niego samego / *odeń samego.
himself / from_him self
'Jan used one of his own citations.'

b. Anna zapytała Jana, czy
Anna asked Jan whether
ma czekać na niego i
she should wait for him and
jego kolegę / *nań i jego
his colleague / for_him and his
kolegę.
colleague
'Anna asked Jan whether she should wait for him and his colleague.'

c. Dla niego / *dlań Maria pracuje od
for him / for_him Maria works for
roku.
year
'For him, Mary has been working for one year.'

Thus, in spite of categorial, syntactic and semantic parallelism, PPCs and ordinary PPs differ with respect to distribution. Unlike typical PPs, PPCs cannot be used in stressed positions, and behave in this regard like typical clitics in Polish. However, we assume that these distributional properties of PPs and PPCs are subject to independent grammatical constraints on stress and intonation patterns in Polish, and for this reason they will not be handled here.

2.4 Lexical Restrictions

It should be noted that the number of PPCs in Polish is relatively small. On one hand the class of pronouns contracting with Ps is limited to genitive, accusative and, possibly, locative singular masculine third person pronouns, on the other hand, the set of prepositions which are able to contract with pronouns involves a very limited number of elements. Thus not every arbitrary preposition can occur in PPC, even if it does combine with genitive or accusative pronouns (cf. (9) and (10)).[11]

(9) Ps combining with genitive pronouns

a. koło niego 'close to him' → *kołoń
b. naprzeciw niego 'opposite to him' → *naprzeciweń
c. podczas niego 'during him' → *podczaseń

[10]For more details on the distribution of weak personal pronouns in Polish see (Kupść, 2000).

[11]Note that the corresponding German *da-*, *wo-* and *hier-* expressions are a semiproductive phenomenon as well. Thus not all German prepositions can combine with *da-*, *wo-* and *hier-* (cf. the examples in (i)).

(i) (a) während 'during' → *wo- / *da- / *hierwährend
(b) mittels 'via' → *wo- / *da- / *hiermittels
(c) zwecks 'in order to' → *wo- / *da- / *hierzwecks

(10) Ps combining with accusative pronouns

a. między nich 'between them' → *międzyń

b. ponad niego 'above him' → *ponadeń

c. poprzez niego 'through him' → *poprzezeń

According to dictionaries of contemporary Polish and considering corpus data, one can claim that only the oldest primary prepositions can contract with pronouns (cf. (11)).

(11) a. bez niego 'without him' → bezeń

b. dla niego 'for him' → dlań

c. do niego 'to him' → doń

d. o niego 'about him' → oń

e. od niego 'from him' → odeń

f. na niego 'on him' → nań

g. nad niego 'above him' → nadeń

h. pod niego 'under him' → podeń

i. przed niego 'in front of him' → przedeń

j. przez niego 'owing him' → przezeń

k. w niego 'in him' → weń

l. z niego 'from him' → zeń

However, as the example in (12) shows, not even all of the oldest primary prepositions undergo this process.

(12) u niego 'by him' → *uń

In view of this fact, it is very difficult to develop a syntactic account for PPC. A lexicon-based treatment seems to be a more plausible and natural option.

3 The Analysis

Based on our discussion of the empirical properties of PPCs, we treat PPCs as morphological units and not as the result of any postlexical processes. Furthermore, we claim that PPCs have the categorial status of a preposition and are valence-saturated but have an internal argument. These generalizations can easily be expressed in the paradigm of HPSG in the tradition of (Pollard and Sag, 1994).

Our analysis of Polish PPC, which will be provided in this section, corresponds to many other constraint-based approaches to similar phenomena. It is worth mentioning the lexicalist treatment of the contraction of French prepositional forms *à* and *de* and the definite article *le* in (Abeillé et al., 2003). Each contraction is analyzed as a lexical item and not as a result of any postlexical operations. (Baronian, 2003) treats these expressions as prepositions with definiteness, number and gender features. (Winhart, 1997) proposes a similar HPSG-based approach to the preposition-determiner contraction in German.[12] (Bender and Sag, 2000) proposes a lexicalist analysis of contracted auxiliaries in English, providing a lexical rule licensing lexical entries that combine information provided by pronouns and auxiliaries. Even expressions traditionally considered to be clitics have often been analyzed lexically, see (Abeillé et al., 1998) and (Miller and Sag, 1997) for a treatment of French pronominal clitics as lexical pronominal affixes attached to inflected verbs. (Monachesi, 1999) also analyzes Italian pronominal clitics as morphological affixes. (Borsley, 1999) provides a lexicon-based analysis of Polish complex verbs. (Kupść and Przepiórkowski, 2002) argues for a morphosyntactic account of Polish verbal negation.

For Polish PPC, we propose a lexicalist analysis as well. Several possibilities could be considered: (1) in a hierarchical lexicon approach, extending the ontology by providing an appropriate lexical sort for each PPC that would then be subject to further constraints, (2) assuming a separate lexical entry for each PPC, (3) specifying all possible distributions in the lexical entry of each preposition, including the possible distribution within a PPC, (4) underspecifying phonological and valence information in lexical entries of Ps and formulating general lexical constraints licensing both PPCs as well as ordinary PPs. Since the third solution has the most general character and seems most elegant, we will assume it here.

We propose thus to assume an underspeficied lexical entry for each preposition being able to contract with pronouns, i.e., *bez* 'without', *dla* 'for', *do* 'to', *o* 'about', *od* 'from', *na* 'on', *nad* 'above', *po* 'for',

[12] In transformational grammar frameworks, the preposition-determiner contraction in German has been analyzed in terms of a Head Adjunction operation (cf. (van Riemsdijk, 1998)).

pod 'under', *przed* 'in front of', *przez* 'owing', *w* 'in', *z* 'from' (cf. the preliminary lexical entry for the preposition *na* 'on' in Figure 1).

$$\begin{bmatrix} \textit{word} \\ \text{MORPH } \langle[\text{PHON } \langle \textit{na}\rangle], \ldots\rangle \\ \text{ARG-ST } \langle[\text{LOC} \mid \text{CAT} \mid \text{HEAD } \textit{noun}]\rangle \\ \text{SYNS} \mid \text{LOC} \mid \text{CAT} \mid \text{HEAD} \begin{bmatrix}\textit{prep}\\ \text{PFORM } \textit{na}\end{bmatrix}\end{bmatrix}$$

Figure 1: The relevant part of the preliminary lexical entry of the preposition *na* 'on'

According to the architecture of linguistic signs and, more precisely, words in terms of (Pollard and Sag, 1994), two different representation levels describe the argument structure of a word and its valence properties. The argument structure of a word is described via the ARG(UMENT)-ST(RUCTURE) attribute taking a list of (*synsem*) objects as its value, while the valence of a word is described via the VAL(ENCE) feature, whose value contains three valence lists: the SP(ECIFIE)R list, the SUBJ(ECT) list and the COMP(LEMENT)S list. Specifying the relationship between the values of these features, one can make generalizations about selectional properties of lexical signs. The ARG-ST value of prepositions is usually assumed to be identical with the COMPS value, while their SUBJ and SPR values are empty lists.

Thus the lexical entry in Figure 1 merely provides information that *na* 'on' can take only one nominal argument. No information about the selection requirements is available here. The syntactic selection properties of *na* 'on' and all other prepositions that are able to contract with pronouns are licensed by the implicational lexical constraint provided in Figure 2.

The antecedent of the lexical constraint in Figure 2 is an underspecified lexical entry for each preposition that can occur in PPC. The P(REPOSITIONAL)FORM attribute, traditionally assumed in HPSG in descriptions of prepositions (cf. (Pollard and Sag, 1994)), that takes names of particular prepositions as its value, makes it possible to identify particular prepositions unambiguously. Note that the argument object is specified to be a genitive or accusative masculine singular third person pronoun.

$$\begin{bmatrix} \textit{word} \\ \text{ARG-ST} \left\langle \begin{bmatrix} \text{LOC} \begin{bmatrix} \text{CAT} \mid \text{CASE } \textit{gen} \vee \textit{acc} \\ \text{CONT} \begin{bmatrix} \textit{ppro} \\ \text{NUMBER } \textit{sing} \\ \text{GENDER } \textit{masc} \\ \text{PERSON } \textit{3rd} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \\ \text{SYNS} \mid \text{LOC} \mid \text{CAT} \mid \text{HEAD} \begin{bmatrix} \textit{prep} \\ \text{PFORM } \textit{dla} \vee \textit{do} \vee \textit{o} \vee \textit{od} \vee \textit{na} \vee \textit{nad} \\ \vee\, \textit{po} \vee \textit{pod} \vee \textit{przed} \vee \textit{przez} \vee \textit{w} \vee \textit{z} \end{bmatrix} \end{bmatrix} \longrightarrow$$

$$\left(\begin{bmatrix} \textit{word} \\ \text{MORPH} \left\langle \textit{morph} \right\rangle \\ \text{ARG-ST } \boxed{1} \\ \text{SYNS} \mid \text{LOC} \mid \text{CAT} \mid \text{VAL} \mid \text{COMPS } \boxed{1} \end{bmatrix} \vee \begin{bmatrix} \textit{word} \\ \text{MORPH} \left\langle \textit{morph}, \left[\text{PHON} \left\langle \textit{ń} \right\rangle\right] \right\rangle \\ \text{SYNS} \mid \text{LOC} \mid \text{CAT} \mid \text{VAL} \mid \text{COMPS} \langle\rangle \end{bmatrix} \right)$$

Figure 2: The lexical constraint for licensing PPC

The first disjunct in the consequence describes prepositions with unrealized complements, licensing ordinary Ps that combine with NP complements. The second disjunct provides a description of prepositional objects with zero valency (see the empty COMPS list), but with an unrealized internal argument (see the ARG-ST value in the antecedent), and licenses PPCs.

The grammar of PPC must also ensure the right phonological form of each PPC. For these purposes, we adopt the approach of (Riehemann, 1998), providing however some modifications. In her hierarchical lexicon-based approach, (Riehemann, 1998) proposes that morphologically complex words, which PPCs definitely are, have a feature MORPH(OLOGICAL)-B(ASE), taking list of signs as its value.[13] We propose to assume a list-valued MORPH(OLOGICAL FORM) feature for each word. The value of the MORPH feature of morphologically simple words is a singleton list, while the value of the MORPH feature of morphologically complex words is a list of *morph* objects.[14]

The lexical entry of the preposition *na* 'on' in Figure 1 involves a partially underspecified MORPH value. While the first element of the MORPH list is specified to be a *morph* object whose phonological form is the string *na*, no specification of possible further elements of that list is provided. The principle in Figure 2 ensures that the MORPH value of ordinary Ps is a singleton list and the MORPH value of

[13]For an alternative HPSG approach to morphology see (Reinhard, 2001) in the context of the analysis of deverbal compounds in German.

[14]We assume that the type hierarchy under the type *sign* consists *phrase*, *word* and *morph*.

PPCs is a list of two *morph* objects. The phonological form of the second *morph* object is always the string *ń*.

The phonological form of PPCs is licensed by the general principle provided in Figure 3 describing phonological form of words.

$$\textit{word} \longrightarrow \begin{bmatrix} \text{PHON } \texttt{F}(\boxed{1} \oplus \ldots \oplus \boxed{n}) \\ \text{MORPH} \left\langle \begin{bmatrix}\text{PHON } \boxed{1}\end{bmatrix}, \ldots, \begin{bmatrix}\text{PHON } \boxed{n}\end{bmatrix} \right\rangle \end{bmatrix}$$

Figure 3: The constraint describing phonologial form of words

According to the principle in Figure 3, the PHON value of words (and thus PPCs) is assumed to be the result of applying a function F to the concatenation of the PHON values of the word's all morphological components. The definition of the F function ensures the right phonological form of words according to phonological constraints for a given language. In the case of Polish PPCs, the output of the F function is the phonological form of a given preposition combined with the phonological form of *-ń*, if that preposition ends in a vowel, and the phonological form of that preposition combined with a phonetic variant of *-ń*, i.e., *-eń*, if that preposition ends in a consonant. Since phonology is not the main subject of this paper, a detailed phonological representation of the contracted forms will not be provided here.[15]

PPC, as most PPs, can act in the sentence either as a complement or as an adjunct, which presupposes different syntactic and semantic properties. We thus postulate a general lexical principle taking a syntactically and semantically underspecified description of prepositions as antecedent and licensing modifying and non-modifying PPs and thus modifying and non-modifying PPCs (cf. Figure 4).

Note that we adopt here the semantic representation of verbs in Davidsonian style (cf. (Davidson, 1967)), i.e., we introduce an event variable into the semantic representation of verbs and assume an architecture of the CONTENT value of verbs analogous to that of nouns.[16] Thus the CONTENT value of a verb is an object containing both an (event) index and a semantic restriction of this index (cf. Figure 5).

[15]For the treatment of phonology in HPSG see (Höhle, 1999).

[16]Cf. (Van Eynde, 1998) or (Sag and Wasow, 1999) for a similar approach to the representation of the verbal semantics in HPSG.

$$\begin{bmatrix} \textit{word} \\ \text{ARG-ST} \left\langle \begin{bmatrix}\text{LOC} \mid \text{CONT} \mid \textit{nom-obj}\end{bmatrix} \right\rangle \\ \text{SYNS} \mid \text{LOC} \mid \text{CAT} \mid \text{HEAD } \textit{prep} \end{bmatrix} \longrightarrow$$

$$\left(\begin{bmatrix} \textit{word} \\ \text{ARG-ST} \left\langle \begin{bmatrix}\text{LOC} \mid \text{CONT } \boxed{1}\end{bmatrix} \right\rangle \\ \text{SYNS} \mid \text{LOC} \begin{bmatrix} \text{CAT} \mid \text{MOD } \textit{none} \\ \text{CONT } \boxed{1} \end{bmatrix} \end{bmatrix} \lor \begin{bmatrix} \textit{word} \\ \text{ARG-ST} \left\langle \begin{bmatrix}\text{LOC} \mid \text{CONT} \mid \text{INDEX } \boxed{2}\end{bmatrix} \right\rangle \\ \text{SYNS} \mid \text{LOC} \begin{bmatrix} \text{CAT} \mid \text{MOD} \begin{bmatrix} \text{LOC} \mid \text{CONT} \begin{bmatrix} \text{INDEX } \boxed{3} \\ \text{RESTR } \boxed{4} \end{bmatrix} \end{bmatrix} \\ \text{CONT} \begin{bmatrix} \text{INDEX } \boxed{3} \\ \text{RESTR} \left\{ \begin{bmatrix} \textit{sem-rel} \\ \text{ARG1 } \boxed{3} \\ \text{ARG2 } \boxed{2} \end{bmatrix} \right\} \cup \boxed{4} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right)$$

Figure 4: The lexical constraint for licensing modifying and non-modifying PPs

$$\begin{bmatrix} \textit{content} \\ \text{INDEX } \boxed{1} \\ \text{RESTR} \left\{ \begin{bmatrix} \textit{psoa} \\ \text{NUCL} \begin{bmatrix} \textit{relation} \\ \text{INST } \boxed{1} \\ \text{ARG1 } \textit{index} \\ \ldots \\ \text{ARG}n \ \textit{index} \end{bmatrix} \end{bmatrix} \right\} \end{bmatrix}$$

Figure 5: The *content* structure of verbs in Davidsonian style

Therefore, the consequence of the principle in Figure 4 consists of two disjuncts, the first of which describes complement PPs and the second PPs modifying NPs and VPs. The semantic representation of complement PPs is, according to traditional analyses, assumed to be identical with the semantic representation of the NPs selected by Ps (cf. the tag $\boxed{1}$ in the first disjunct). The semantic representation of adjunct PPs is an object whose INDEX attribute's value is identical with the INDEX value of the modified object, and whose RESTRICTION feature's value is a union of its own RESTRICTION set and the RESTRICTION set of the modified object.

Note that the *index* of the argument (see the identity of the CONTENT values, which contain an *index*, in the first disjunct and the tag $\boxed{2}$ in the second disjunct) is present in the semantic representation of each disjunct, and thus control phenomena such as those exemplified in (7) can easily be captured.

While the principle in Figure 2 accounts for a limited number of Ps taking genitive and accusative third person pronouns as their arguments, the principle in Figure 6 covers all other prepositions which take arbitrary arguments and combine with them according to constraints on phrase structure.

$$\begin{bmatrix} \textit{word} \\ \text{ARG-ST} \left\langle \begin{array}{l} \begin{bmatrix} \text{LOC} \begin{bmatrix} \text{CAT} \mid \text{HEAD} \mid \text{CASE } \textit{nom} \vee \textit{dat} \vee \textit{inst} \vee \textit{loc} \\ \text{CONT } \textit{ppro} \end{bmatrix} \end{bmatrix} \\ \vee \begin{bmatrix} \text{LOC} \mid \text{CONT} \mid \textit{npro} \vee \textit{ana} \end{bmatrix} \end{array} \right\rangle \\ \text{SYNS} \mid \text{LOC} \mid \text{CAT} \mid \text{HEAD } \textit{prep} \end{bmatrix} \longrightarrow \begin{bmatrix} \textit{word} \\ \text{ARG-ST } \boxed{1} \\ \text{SYNS} \mid \text{LOC} \mid \text{CAT} \mid \text{VAL} \mid \text{COMPS } \boxed{1} \end{bmatrix}$$

Figure 6: The lexical constraint describing prepositions selecting arbitrary arguments

Finally, it should be ensured that prepositions, when combining with third person pronouns, combine with postprepositional pronouns exclusively. According to (Saloni, 1981), as well as many grammars of Polish, Polish third person personal pronouns inflect besides case (nominative, genitive, dative, accusative, instrumental and locative) number (singular and plural) and accentability (yes or not) for postprepositionality (yes or not). Thereby the grammatical category of postprepositionality relates to structural circumstances rather than to prosodic (cf. the examples in (13)).

(13) a. Źle się to skończyło dla niej i niego / *jego.
bad RM it ended for her and him$_{pp}$ / him$_{non\text{-}pp}$
'It came to a bad end for her and him.'

b. Ta wiadomość jest zła zarówno dla nas jak i niego / *jego samego.
this message is bad both for us like and him$_{pp}$ / him$_{non\text{-}pp}$ self
'This message is bad both for us and for him self.'

Thus the postprepositionality value of all personal pronouns in (13) must be positive, since all of them are syntactically selected by prepositions as a part of a coordination. Although pronouns in the second conjunct follow a conjunction and are thus not adjacent to a preposition, they bear the positive postprepositionality value.[17]

We propose thus to assume the description of third person pronouns in Polish as in Figure 7.

$$\begin{bmatrix} \textit{word} \\ \text{ARG-ST} \langle\rangle \\ \text{SYNS} \mid \text{LOC} \begin{bmatrix} \text{CAT} \mid \text{HEAD} \begin{bmatrix} \textit{noun} \\ \text{PP} + \end{bmatrix} \\ \text{CONT} \begin{bmatrix} \textit{ppro} \\ \text{INDEX} \mid \text{PERSON } \textit{3rd} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Figure 7: A simplified description of postprepositional third person pronouns

In order to be able to identify postprepositional pronouns, we propose to introduce a new feature appropriate for the sort *noun*, i.e., the feature P(OST)P(REPOSITIONALITY) taking boolean values. The PP value of non-postprepositional pronouns such as *jego* 'his', *je* 'it' or *ją* 'her' is $-$, while the PP value of postprepositional pronouns such as *niego* 'his', *nie* 'it' or *nią* 'her' is $+$. Non-pronominal nouns such as *Piotr* 'Piotr' or książka 'book', as well as nominative, dative, instrumental and locative personal pronouns and anaphora have an underspecified PP value.

According to this modification of the appropriateness conditions for the sort *noun*, the description of postprepositional pronouns in Figure 7 involves the PP feature taking the value $+$.

Given this, a constraint can be formulated that ensures that if a preposition selects a pronoun, the pronoun must bear a positive postprepositionality feature (cf. Figure 8).

$$\begin{bmatrix} \textit{word} \\ \text{SYNS} \begin{bmatrix} \text{LOC} \mid \text{CAT} \mid \text{HEAD } \textit{prep} \end{bmatrix} \end{bmatrix} \longrightarrow \begin{bmatrix} \text{ARG-ST} \left\langle \begin{bmatrix} \text{LOC} \mid \text{CAT} \mid \text{PP} + \end{bmatrix} \right\rangle \end{bmatrix}$$

Figure 8: The constraint ensuring the right PP value of objects selected by prepositions

The principle in Figure 9 guarantees the negative value of the PP attribute if a pronoun (as well as all

[17]In (13) I use the abbreviation *pp* to mark pronouns with the positive postprepositionality value and the abbreviation *non-pp* to mark pronouns with the negative postprepositionality value.

other objects with the PP attribute) is not selected by a preposition.

$$\begin{bmatrix} \textit{word} \\ \text{SYNS}\left[\text{LOC} \mid \text{CAT} \mid \text{HEAD} \neg \textit{prep}\right]\end{bmatrix} \longrightarrow \left[\text{ARG-ST}\left\langle\left[\text{LOC} \mid \text{CAT} \mid \text{PP} -\right]\right\rangle\right]$$

Figure 9: The constraint ensuring the right PP value of objects selected by non-prepositional objects

4 Summary and Outlook

In this paper, we have discussed preposition-pronoun contraction in Polish. A range of data has been examined with respect to the prosodic, categorial, syntactic and semantic properties. On the basis of empirical observation, the generalization has been made that a PPC is a morphological unit which has prepositional status.

An HPSG-based lexicalist approach has been provided which accounts for both PPC and all other types of prepositions and prepositional phrases in Polish.

In future work, a corpus-based study on third person pronouns will be carried out in order to determine the number of pronominal forms exactly which can occur within PPCs in contemporary Polish.

It may also be interesting to examine whether the analysis proposed here for Polish data can be applied to corresponding data in other languages, e.g., to *da- / wo- / hier-* expressions in German.

5 Acknowledgments

We would like to thank Adam Przepiórkowski, Frank Richter and the anonymous reviewers of The Second ACL-SIGSEM Workshop on The Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications in Colchester, United Kingdom, for their helpful comments. We are also grateful to Janina Radó and Guthrun Love for their help with the proofreading in English.

References

Anne Abeillé, Danèle Godard, and Ivan A. Sag. 1998. Two Kinds of Composition in French Complex Predicates. In Erhard Hinrichs, Andreas Kathol, and Tsuneko Nakazawa, editors, *Complex Predicates in Nonderivational Syntax*, pages 1–41, New York. Academic Press.

Anne Abeillé, Olivier Bonami, Danièle Gorard, and Jesse Tseng. 2003. The Syntax of French *à* and *de*: An HPSG Analysis. In *Proceedings of the ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*, pages 133–144, Toulouse, France. IRIT.

Piotr Bąk. 2004. *Gramatyka języka polskiego [Grammar of Polish].* Wiedza Powszechna, Warszawa, xii edition.

Mirosław Bańko. 2000. *Inny słownik języka polskiego [Different Polish Dictionary].* Wydawnictwo Naukowe PWN, Warszawa.

Luc V. Baronian. 2003. Preposition Contractions in Quebec French. In *Proceedings of the ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*, pages 20–31, Toulouse, France. IRIT.

Emily M. Bender and Ivan A. Sag. 2000. Incorporating Contracted Auxiliaries in English. In Ronnie Cann, Claire Grover, and Philip Miller, editors, *Grammatical Interfaces in HPSG*, pages 17–32. CSLI, Stanford.

Robert D. Borsley. 1999. Auxiliaries, Verbs and Complementizers in Polish. In Robert D. Borsley and Adam Przepiórkowski, editors, *Slavic in HPSG*, pages 29–59. CSLI Publications, Stanford.

Donald Davidson. 1967. The Logical Form of Action Sentences. In Nicholas Rescher, editor, *The Logic of Decision and Action*. University of Pittsburgh Press, Pittsburgh.

Witold Doroszewski and Bolesław Wieczorkiewicz. 1972. *Gramatyka opisowa języka polskiego z ćwiczeniami [A Descriptive Grammar of Polish with Exercises]*, volume II: Fleksja. Składnia [Inflection. Syntax.]. Państwowe Zakłady Wydawnictw Szkolnych, Warszawa.

Stanisław Dubisz. 2003. *Uniwersalny słownik języka polskiego [The Universal Polish Dictionary].* Wydawnictwo Naukowe PWN, Warszawa.

Duden IV. 1984. *Grammatik der deutschen Gegenwartssprache [Grammar of Contemporary German].* Dudenverlag, Mannheim.

Ulrich Engel. 1996. *Deutsche Grammatik [German Grammar].* Julius Groos Verlag, 3rd edition.

Norbert Fries. 1988. *Präpositionen und Präpositionalphrasen im Deutschen und im Neugriechischen [Prepositions and Prepositional Phrases in German and Modern Greek]*. Number 208 in Linguistische Arbeiten. Max Niemeyer Verlag, Tübingen.

Renata Grzegorczykowa, Roman Laskowski, and Henryk Wróbel. 1984. *Gramatyka współczesnego języka polskiego: Morfologia [The Grammar of Contemporary Polish: Morphology]*. Państwowe Wydawnictwo Naukowe, Warszawa.

Gerhard Helbig and Joachim Buscha. 1993. *Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht [German Grammar. A Handbook for Second Language]*. Langenscheidt, Verlag Enzyklopädie, Leipzig, Berlin, München, Wien, Zürich, New York.

Tilman N. Höhle. 1999. An Architecture for Phonology. In Robert D. Borsley and Adam Przepiórkowski, editors, *Slavic in Head-Driven Phrase Structure Grammar*, Studies in Constraint-Based Lexicalism, pages 61–90. CSLI Publications, Stanford.

Anna Kupść and Adam Przepiórkowski. 2002. Morphological Aspects of Verbal Negation in Polish. In Peter Kosta and Jens Frasek, editors, *Current Approaches to Formal Slavic Linguistics*, pages 337–346, Frankfurt am Main. Peter Lang.

Anna Kupść. 2000. *An HPSG Grammar of Polish Clitics*. Ph.D. thesis, Institute of Computer Science, Polish Academy of Sciences and Paris 7.

Tadeusz Lehr-Splawiński and Roman Kubiński. 1952. *Gramatyka języka polskiego [Grammar of Polish]*. Wydawnictwo Zakładu Narodowego im. Ossolińskich, Wrocław, 6 edition.

Philip H. Miller and Ivan A. Sag. 1997. French Clitic Movement without Clitics or Movement. *Natural Language and Linguistic Theory*, 15:573–639.

Paola Monachesi. 1999. *A Lexical Approach to Italian Clitization*. CSLI Publications, Stanford.

Alicja Nagórko. 1996. *Zarys gramatyki polskiej [An Outline of Polish Grammar]*. Wydawnictwo Naukowe PWN, Warszawa.

Carl J. Pollard and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. The University of Chicago Press, Chicago.

Adam Przepiórkowski. 2004. *The IPI PAN Corpus. Preliminary Version*. Institute of Computer Science PAS, Warsaw.

Sabine Reinhard. 2001. *Deverbale Komposita an der Morphologie-Syntax-Semantik-Schnittstelle: ein HPSG-Ansatz [Deverbal Compounds at the Morphology-Syntax-Semantics Interface: an HPSG Approach]*. Phil. dissertation, Eberhard-Karls-Universität Tübingen.

Susanne Riehemann. 1998. Type-Based Derivational Morphology. *Journal of Comparative Germanic Linguistics*, 2:49–77.

Ivan A. Sag and Thomas Wasow. 1999. *Syntactic Theory: A Formal Introduction*. CSLI Publications.

Zygmunt Saloni and Marek Świdziński. 1998. *Składnia współczesnego języka polskiego [The Syntax of Contemporary Polish]*. Wydawnictwo Naukowe PWN, Warszawa, 4 edition.

Zygmunt Saloni. 1981. Uwagi o opisie fleksyjnym tzw. zaimków rzeczownych [Some Remarks on the Inflexional Description of Polish Pronouns]. In *Acta Universitatis Lodziensis*, volume 2 of *Folia Linguistica*, pages 243–253. Uniwersytet Łódzki.

Marek Świdziński. 1992. *Gramatyka formalna języka polskiego [Formal Grammar of Polish]*, volume 349 of *Rozprawy Uniwersytetu Warszawskiego*. Wydawnictwa Uniwersytetu Warszawskiego, Warszawa.

Stanisław Szober. 1969. *Gramatyka języka polskiego [The Grammar of Polish]*. PWN, Warszawa.

Frank Van Eynde. 1998. Tense, Aspect and Negation. In F. Van Eynde and P. Schmidt, editors, *Linguistic Specifications for Typed Feature Structure Formalisms. Studies in Machine Translation and Natural Language Processing*, volume 10, pages 209–280. Office for Official Publications of the European Communities, Luxembourg.

Henk van Riemsdijk. 1998. Head Movement and Adjacency. *Natural Language and Linguistic Theory*, 16:633–678.

Bronisław Wieczorkiewicz, Maria Froelichowa, and Salomea Szlifersztejnowa. 1957. *Gramatyka polska [Polish Grammar]*. Państwowe Zakłady Wydawnictw Szkolnych, Warszawa, 6 edition.

Heike Winhart. 1997. Die Nominalphrase in einem HPSG-Fragment des Deutschen. In Erhard Hinrichs, Detmar Meurers, Frank Richter, Manfred Sailer, and Heike Winhart, editors, *Ein HPSG-Fragment des Deutschen, Teil 1: Theorie*, volume 95 of *Arbeitspapiere des SFB 340*, pages 319–384. Universität Tübingen.

Henryk Wróbel. 2001. *Gramatyka języka polskiego [The Grammar of Polish Language]*. Spółka Wydawnicza "OD NOWA", Kraków.