Formal Description of Natural Languages: An HPSG Grammar of Polish

Leonard Bolc

Institute of Computer Science,
Polish Academy of Sciences, Warsaw

1 Introduction

In this paper we present a Head-driven Phrase Structure Grammar (HPSG) grammar of Polish – the result of one of the few attempts (e.g., [Szp86], [Świ92]) to build formal and computationally tractable grammars of Polish. The choice of the formalism used was motivated by several promising features of the formalism which we will present shortly.

The research concerning HPSG description of Polish started in 1994, when members of the Linguistic Engineering Group of Institute of Computer Science, Warsaw, have undertaken research aimed at the description of the large subset of Polish syntax in the terms of this formalism. At the beginning of the work separate syntactic issues were worked up and some theories based on the fundamental HPSG theory described in [PS94] were formulated. The need of the coherent theory of Polish syntax which can become a foundation of the implementation of a relatively large Polish grammar led to further work aimed at integration of all subtheories. The effect of these efforts is the book "Formalny opis języka polskiego. Teoria i implementacja" (Formal description of Polish. Theory and implementation.) by Adam Przepiórkowski, Anna Kupść, Małgorzata Marciniak and Agnieszka Mykowiecka. This paper is a short presentation of the results of the efforts to describe Polish within HPSG formalism included in that book.

HPSG was developed as a comprehensive linguistic formalism for work on syntax, morphology and semantics, as well as phonology and pragmatics. It is a monostratal theory of language: there are no derivations transforming one grammatical structure into another. Any grammatical structure is well-formed if it simultaneously satisfies all constraints that the grammar imposes. Further, all constraints are local, limited to one structure at a time. HPSG puts emphasis on explicitness and precision, its linguistic analyses are couched in a mathematical formalism with well-defined syntax and model-theoretic semantics. Because of this explicitness and formality, HPSG has become one of the most popular linguistic formalisms in computational linguistic applications and this is one of the most important reasons why we have chosen it in our work.

HPSG is a linguistic formalism, i.e., a set of formal tools for formalising linguistic analyses of various phenomena, but it is also a linguistic theory, i.e., a collection of analyses of various phenomena described using this formalism. In this work we accept the main ideas of the formalism but at the same time we

D. Hutter, W. Stephan (Eds.): Mechanizing Mathematical Reasoning, LNAI 2605, pp. 297–320, 2005.

introduce some changes in the theory itself and in ways of representing particular aspects of the linguistic constructs.

HPSG grammars consist of a signature and a theory proper. The theory is a set of constraints that all objects in the model must simultaneously satisfy. The signature defines what types of objects there are (e.g., verbs, nouns, cases, genders) and what features they may have (e.g., verbs have person but not case, nouns have case, genders are atomic objects, i.e., do not have any features). In particular all linguistic expressions are represented by objects of the type *sign* having two subtypes: *phrase* and *word*.

The next part of any HPSG theory is a set of constraints. The most famous HPSG constraint is the Head Feature Principle, a version of which is given in (1).

(1) $$\textit{phrase} \rightarrow \begin{bmatrix} \textsc{synsem|local|cat|head} & \boxed{1} \\ \textsc{head-dtr|synsem|local|cat|head} & \boxed{1} \end{bmatrix}$$

Head Feature Principle is an implicational constraint: every object that is characterised by the left-hand side of '→' must also be characterised by the right-hand side. In this particular case, every object of type *phrase* must be such that the value of its SYNSEM|LOCAL|CAT|HEAD attribute is also the value of the SYNSEM|LOCAL|CAT|HEAD attribute of its head daughter. The tag '1' is just a variable used for indicating equality between paths.

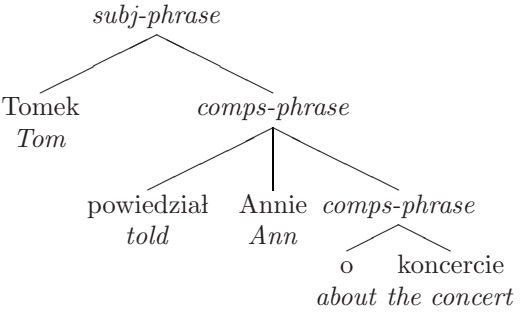
We will not introduce here the HPSG theory itself, the reader is referred to [PS94]. We will focus on presenting new elements of the theory and their interpretation.

2 Modifications of the Standard Theory

2.1 "Flat" Phrase Structure

According to the generally accepted assumption, in HPSG (and in other generative formalisms) the head element (e.g., verb *powiedział 'told'*) takes first its complements (e.g., noun *Annie 'Ann'* and prepositional phrase *o koncercie 'about the concert'*) forming almost saturated phrase, e.g., verb phrase *powiedział Annie o koncercie 'told Ann about the concert'*. Then this phrase takes the subject and forms a saturated phrase, i.e., a structure with empty valence lists SUBJ and COMPS, e.g., a clause *Tomek powiedział Annie o koncercie* (see (2) below).

(2)

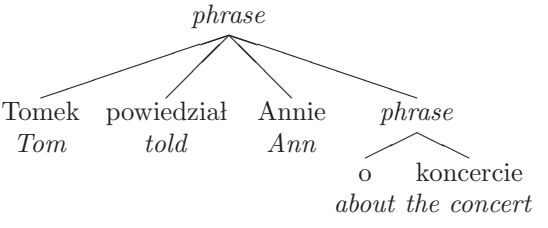


Arguments for two stages' phrase construction, that is separate realisation of subject and complements, are not sufficiently convincing for Polish. There is no place here for a detailed discussion of the subject (such a discussion can be found in [PKMM01]), so we present only examples (3) showing that in Polish there are no order rules supporting the distinction between the subject and complements. Sentences in which the subject is realised 'closer' to a verb than its complement are quite frequent in Polish.

(3) a. Pwiedział Tomek Annie o koncercie.
b. O koncercie powiedział Annie Tomek.
c. Annie Tomek powiedział o koncercie.

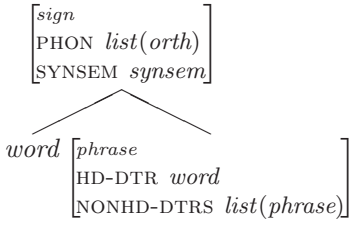
Consequently, we assume that all arguments of a head are syntactically realized at the same level of phrase structure. As a result, phrase structures are flat, as in (4) below, and there is no need for the distinction between types *subj-phrase* and *comps-phrase*.

(4)



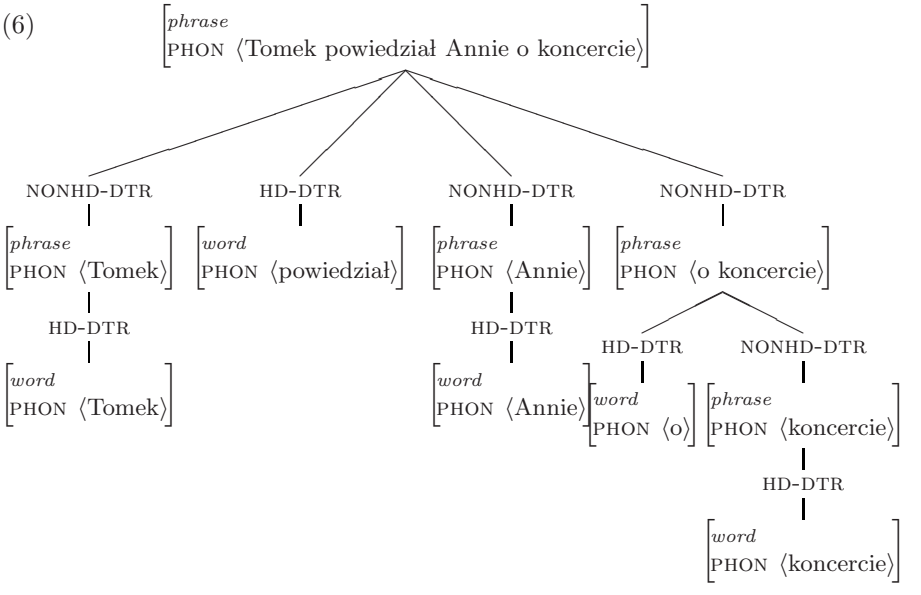
The next assumption made by us about the construction of phrases is the constraint imposed on the types of phrase elements. It states that the head element of the phrase should be a *word* (not a *phrase*) while the elements of the NONHD-DTRS list should be *phrase*s (not *word*s). This assumption is formalised in the hierarchy of the *sign* type below, (5).

(5)



The above assumption allows for eliminating redundant parses which differ only in treating a particular input element as a word or as a (one word) phrase. At the same time, it imposes the simultaneous realisation of all the arguments. For example, the only parse tree for the sentence *Tomek powiedział Annie o koncercie* will be the that given in (6).

(6)



The changes just introduced allow for the simpler formulation of the Valence Principle. As all head's arguments have to be satisfied simultaneously, the value of the phrases' VALENCE attribute has to be an empty list. Thus, it turns out that for phrases this attribute is not necessary at all. So, we introduce it only for objects of the type *word* (to do so, we must change the place of introducing this attribute to the highest level within the *word* structure). Since we proposed not to distinguish subjects and complements at the phrase structure level, the division of a VALENCE list into SUBJ and COMPS attributes is no longer necessary – the VALENCE attribute simply has a list of *synsem*s as its value:

(7) $\begin{bmatrix} word \\ \text{VALENCE } list(synsem) \end{bmatrix}$

The above changes make it possible to express the Valence Principle in the following way:

(8) **Valence Principle**

$$phrase \rightarrow \begin{bmatrix} \text{HD-DTR}|\text{VAL } \boxed{1} \\ \text{NONHD-DTRS } \boxed{2} \end{bmatrix} \wedge \texttt{synsems-signs(}\boxed{1}\texttt{,}\boxed{2}\texttt{).}$$

(9) `synsems-signs(<>,<>).`

$$\texttt{synsems-signs(<}\boxed{1}\texttt{|}\boxed{2}\texttt{>,<}\boxed{1'}\texttt{|}\boxed{2'}\texttt{>)} \quad \overset{\forall}{\Longleftarrow}$$

$$\boxed{1'} = \begin{bmatrix} sign \\ \text{SYNSEM } \boxed{1} \end{bmatrix}$$

$$\wedge \texttt{ synsems-signs(}\boxed{2}\texttt{,}\boxed{2'}\texttt{).}$$

In particular, the Valence Principle requires that the head element has the VALENCE attribute, so it automatically excludes phrases as head elements.

2.2 The Correspondence Between ARG-ST and VALENCE

In this subsection we will describe the relation between VALENCE and ARG-ST attributes. First, although we think that the distinction between subjects and complements plays no role in describing the structure of sentences, it is important for describing some language phenomena, such as agreement, case assignment or binding theory. Since these phenomena are accounted for with the help of the ARG-ST attribute, we re-introduce the subject/complements distinction at the level of ARG-ST list and posit that values of this argument be structures of the following *arg-st* type:

(10)
$$\begin{bmatrix} \textit{arg-st} \\ \text{SUBJ } \textit{list}(\textit{synsem}) \\ \text{ARGS } \textit{list}(\textit{synsem}) \end{bmatrix}$$

The next change concerning argument structure description concerns defining the ARG-ST attribute within the *head* structure. Consequently, the ARG-ST attribute is now defined not only for *words* but also for *phrases* (a detailed discussion concerning this problem may be found in [Prz01]).

(11)
$$\begin{bmatrix} \textit{category} \\ \text{HEAD} \begin{bmatrix} \textit{head} \\ \text{ARG-ST} \begin{bmatrix} \textit{arg-st} \\ \text{SUBJ } \textit{list}(\textit{synsem}) \\ \text{ARGS } \textit{list}(\textit{synsem}) \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Attributes ARG-ST (SUBJ and ARGS) and VALENCE include similar but not necessarily the same elements. On the VALENCE list we put those elements from the ARG-ST|SUBJ and ARG-ST|COMPS lists which are realised in the local syntactic tree, while on the ARG-ST lists all predicate arguments, even those not syntactically realised, are present. These non realised arguments can be of three following kinds:

- dummy subject of personal verb forms (***pro***),
- subject of non personal verb forms (e.g., inifinitive *ogolić się* in (12) or participle *myśląc* w (13)).
  (12) Kazał Tomkowi się ogolić. He told Tom to shave himself.
  (13) Jadła myśląc o swojej przyszłości. She ate thinking about her future.
- arguments of verbs located 'lower' in the syntactic tree realised 'higher' in the syntactic structure, e.g., a complement of the verb *zaprosić* in (14) realised 'higher' as a interrogative pronoun or a subject realised in (15) as a relative pronoun *który*. Such non-locally realised arguments are called *gaps*.
  (14) Kogo chciałeś, żebym zaprosił __?
  Who you wanted that me invited __
  'Who did you want me to invite?'
  (15) ... facet, który chciałam, żeby __ przyszedł.
  ... guy who I wanted that __ came
  '... a guy whom I wanted to come'

To allow for nonlocal arguments we introduce (after [MS97], [Sag97] and [BMS01]) two subtypes of the *synsem* type: *canonical-synsem* (*canon-ss*), representing arguments which are locally realised, and *noncanonical-synsem* (*non-canon-ss*), representing non-realised arguments.

(16)
synsem
canon-ss *noncanon-ss*

We require that values of the SYNSEM attribute be of the *canon-synsem* type, see (17). This will cause all VALENCE elements to have their SYNSEM values of the type *canon-synsem*, so objects of the *noncanon-synsem* can appear only on the ARG-ST lists.

(17)
$$\begin{bmatrix} \textit{sign} & \\ \text{PHON} & \textit{list}(\textit{orth}) \\ \text{SYNSEM} & \textit{canon-ss} \end{bmatrix}$$

A revised version of Argument Structure Principle is given below[1]:

(19) **Argument Structure Principle**

$$\textit{word} \rightarrow \begin{bmatrix} \text{SYNSEM}|\text{LOCAL}|\text{CAT}|\text{HEAD}|\text{ARG-ST} & \begin{bmatrix} \textit{arg-st} & \\ \text{SUBJ} & \boxed{1} \\ \text{ARGS} & \boxed{2} \end{bmatrix} \\ \text{VALENCE} & \boxed{3} \end{bmatrix}$$
$$\wedge\ \boxed{1} \oplus \boxed{2} = \boxed{3} \bigcirc \textit{list}(\textit{noncanon-ss}).$$

In the following section we will introduce the notion of raising elements. The distinction between raised and non-raised elements make the actual *synsem* hierarchy a little more complicated (see (20)), but the formulation of the Argument Structure Principle remains unchanged.

2.3 Noncanonical Arguments

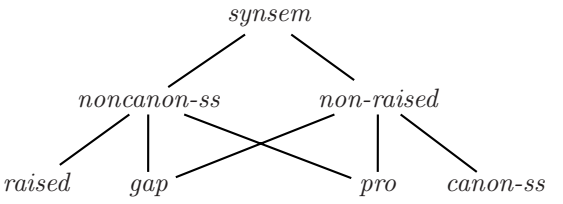
In this section, we will present non-canonical arguments in more detail and we will introduce further refinements of the *synsem* hierarchy, given in (20). We will not present here the detailed discussion on their distribution, which can be found in [PKMM01].

---

[1] "$\oplus$" is an abbreviation for the `append` relation; "$\bigcirc$" is an infix notation of the shuffle relation [Rea92], i.e., `shuffle(`$\boxed{1}$`,`$\boxed{2}$`,`$\boxed{3}$`)` $\equiv \boxed{1} \bigcirc \boxed{2} = \boxed{3}$. The `shuffle` relation is defined as follows:

(18)
```
shuffle(<>,<>,<>).
shuffle(<[1]|[2]>,[3],<[1]|[4]>)  ⇐∀  shuffle([2],[3],[4]).
shuffle([2],<[1]|[3]>,<[1]|[4]>)  ⇐∀  shuffle([2],[3],[4]).
```

(20)



Raised Arguments. We assume that some arguments may be 'raised' higher in the syntactic hierarchy instead of being realised locally; i.e. they are realised as syntactic arguments of the higher verb. An example of such a construction is an infinitival phrase. In the sentence below, arguments of the infinitive verb *dać 'give'* can be realised locally or they can be raised and realised as arguments of the verb *chciał 'wanted'*.

(21) Janek chciał dać Marysi kwiaty.
'John wanted to give Mary flowers.'

Partial motivation for introducing argument raising comes form the phenomenon of Genitive of Negation. If the higher verb (i.e., *chciał*) is negated, the argument *kwiaty* should occur in genitive, not in accusative; compare (22) with (21).

(22) Janek nie chciał dać Marysi kwiatów.

In order to maintain local case assignment principles, we are forced to assume that, in (21), *kwiaty* is in some sense the argument of the verb *chiał*. On the other hand, Genitive of Negation is to some extent optional. In some cases, such arguments may stay in the accusative case (cf. [Prz99,Prz00]). Consequently, we assume that argument raising is optional, i.e., examples like (21) have several parses differing in the placement of infinitival's arguments.

Lexical entries do not specify which arguments are of the *canon-ss* type, i.e., which arguments are realised locally. However, they have to represent the fact that those arguments which are not realised locally have to be raised to a higher level. To represent objects which can be raised, we introduce the *raised – non-raised* distinction within the *synsem* type. All *raised* objects are of *noncanon-ss* subtype while *non-raised* arguments are divided into *canon-ss*, *pro* and *gap* subtypes, (20).

The adequate lexical entry for the verb *chciał* is given below. The `raise-local` function is responsible for raising LOCAL structures only. Subjects of infinitives are never locally realised, so their SUBJ value is of the *noncanon-ss* type.

(23)

$$\left[\begin{array}{l} word \\ \text{PHON } \langle \text{chciał} \rangle \\ \text{SS|L|C|H|ARG-ST } \left\langle \left[\begin{array}{l} arg\text{-}st \\ \text{SUBJ } \langle \left[\begin{array}{l} synsem \\ \text{L|C|H } \left[\begin{array}{l} noun \\ \text{CASE } nom \end{array}\right] \end{array}\right] \rangle \\ \text{ARGS } \langle \left[\begin{array}{l} synsem \\ \text{L|C|H } \left[\begin{array}{l} inf \\ \text{ARG-ST } \left[\begin{array}{l} arg\text{-}st \\ \text{SUBJ } \langle noncanon\text{-}ss \rangle \\ \text{ARGS } list(canon\text{-}ss) \bigcirc \boxed{2} list(raised) \end{array}\right] \end{array}\right] \end{array}\right] \rangle \oplus \boxed{2'} \end{array}\right] \end{array}\right]$$

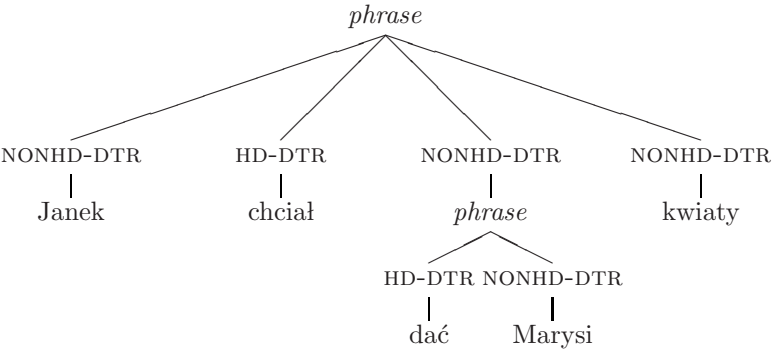
$\wedge$ `raise-local(`$\boxed{2}$`,`$\boxed{2'}$`).`

(24) `raise-local(<>,<>).`

`raise-local(<`$\left[\begin{array}{l} synsem \\ \text{LOCAL } \boxed{1} \end{array}\right]$`|`$\boxed{2}$`>,<`$\left[\begin{array}{l} synsem \\ \text{LOCAL } \boxed{1} \end{array}\right]$`|`$\boxed{3}$`>)` $\overset{\forall}{\Longleftarrow}$ `raise-local(`$\boxed{2}$`,`$\boxed{3}$`).`

If we assume that the dative complement of *dać* is realised locally, while the accusative complement is not, we will achieve the structure given in (25).

(25)



Pro. The other kind of non-canonical argument is a dummy subject, *pro*. In (26), the subject is not realised on the surface, so it can be present only in the ARG-ST|SUBJ list, not in the VALENCE list. Dummy subjects are represented by a special subtype of the *noncanon-ss* type – *pro*[2].

(26) a. Dał Marysi kwiaty.

b.
$$\left[\begin{array}{l} word \\ \text{PHON } \langle \text{dał} \rangle \\ \text{SYNSEM|LOCAL|CAT|HEAD|ARG-ST } \left[\begin{array}{l} arg\text{-}st \\ \text{SUBJ } \langle \boxed{1} pro \rangle \\ \text{ARGS } \langle \boxed{2}\text{NP}[dat], \boxed{3}\text{NP}[acc] \rangle \end{array}\right] \\ \text{VALENCE } \langle \boxed{2}, \boxed{3} \rangle \end{array}\right]$$

[2] We give up here the traditional (in the generative literature) distinction between PRO and *pro* and represent both kinds of empty elements as *pro*.

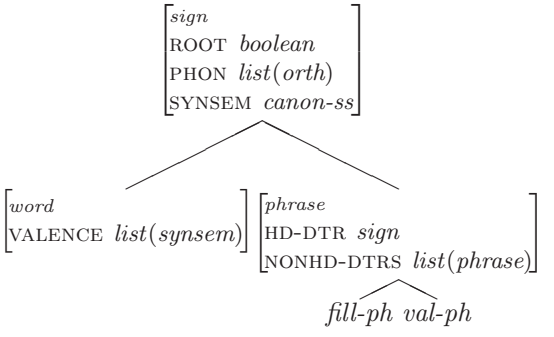
It is not always true, that a non realised subject is represented by the *pro* element. There exist arguments that infinitival subjects in sentences like (27) should not be represented as *pro* but should be treated as being raised. Such an analysis reflects number and gender agreement between *Janek* – the subject of the higher verb and the adjective *miły*.

(27) Janek chciał __ być miły.
John wanted to be nice.

2.4 Types of Phrases

We assume two phrase types: *valence-phrase* and *fill-phrase* (named *val-ph* nad *fill-ph* respectively). The complete hierarchy of *sign* is presented below:

(28)



Valence-Phrase. The basic phrase schema described in 2.1 represents phrases which consist of a head element (a word) and its dependents. This schema can be used to build not only clauses but also phrases with non-verbal head elements, e.g., nominal or prepositional phrases. It is also used for representing phrases with markers (complementizers), which are analysed as heads. To allow for this, we introduce a *marker* subtype of type *head*.

We call all phrases constructed according to this basic schema *valence-phrase* (*val-ph*) and limit the scope of the Valence Principle to this type of phrases only:

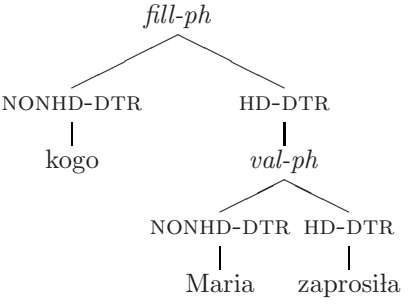
(29) **Valence Principle**

$$\textit{valence-phrase} \rightarrow \begin{bmatrix} \text{HD-DTR}|\text{VAL } \boxed{1} \\ \text{NONHD-DTRS } \boxed{2} \end{bmatrix} \wedge \texttt{synsems-signs}(\boxed{1},\boxed{2}).$$

A phrase of type *val-ph* is always most deeply nested, because its head element should have the VALENCE attribute, so it should be a word. For example, in the phrase in (30), a *val-ph* phrase has to occur inside a *fill-ph* phrase.

(30) a. Mówiłeś, że [kogo Maria zaprosiła]?

b.



Filler-Phrase. Traditionally, *filler-phrases* are used to represent structures within which some locally non-realised elements are finally realised. For example in the phrases below, the elements *kto*, *co*, *komu*, *którego*, *komu* are realised nonlocally.

(31) [Kto co komu] [dał]?
(32) ... facet, [któremu] [chciałaś, żebym dał tę książkę].

The analysis of non-local dependencies follows [BMS01] and it is based on the idea of passing information about non-realised arguments via the NONLOCAL structure. Non-realised elements are locally represented by the special object of the type *gap*, which is the last subtype of *noncanon-ss* that we define here. This type introduces a nonempty value of the SLASH attribute within the NONLOCAL structure, (33).

(33) $$gap \rightarrow \begin{bmatrix} \text{LOCAL } \boxed{1} \\ \text{NONLOC|SLASH } \langle \boxed{1} \rangle \end{bmatrix}$$

For example in the case of *dał*, (31), the NONLOCAL|SLASH value is as follows:

(34) $$dał\text{:} \begin{bmatrix} nonlocal \\ \text{SLASH } \langle \begin{bmatrix} local \\ \text{C|H } \begin{bmatrix} noun \\ \text{CASE } nom \end{bmatrix} \end{bmatrix}, \begin{bmatrix} local \\ \text{C|H } \begin{bmatrix} noun \\ \text{CASE } acc \end{bmatrix} \end{bmatrix}, \begin{bmatrix} local \\ \text{C|H } \begin{bmatrix} noun \\ \text{CASE } dat \end{bmatrix} \end{bmatrix} \rangle \end{bmatrix}$$

We impose the following constraint on the *fill-ph* type[3]:

(35) $$fill\text{-}ph \rightarrow \begin{bmatrix} \text{SYNSEM|NONLOC|SLASH } \langle\rangle \\ \text{HD-DTR } \begin{bmatrix} val\text{-}ph \\ \text{SS|NONLOC|SLASH } \boxed{1} nelist \end{bmatrix} \\ \text{NONHD-DTRS } \boxed{1'} \end{bmatrix} \wedge \texttt{locals-signs(}\boxed{1}\texttt{,}\boxed{1'}\texttt{).}$$

$$\texttt{locals-signs(<>,<>).}$$

$$\texttt{locals-signs(<}\boxed{1}\texttt{|}\boxed{2}\texttt{>,<}\boxed{1'}\texttt{|}\boxed{2'}\texttt{>)} \overset{\forall}{\Longleftarrow}$$

$$\boxed{1'} = \begin{bmatrix} sign \\ \text{SYNSEM|LOCAL } \boxed{1} \end{bmatrix}$$

$$\wedge\ \texttt{locals-signs(}\boxed{2}\texttt{,}\boxed{2'}\texttt{).}$$

[3] We require the SLASH value to be nonempty in order to exclude trivial *fill-ph* phrases with no non-locally realised elements.

2.5 Lexicon

Although it is possible to construct a more sophisticated lexicon structure taking advantage of HPSG type hierarchies, we adopt the simplest solution and define the lexicon as a set of lexical entries. A slight modification of the standard approach consists in introducing a difference between lexical entries (of type *entry*) and syntactic words (of type *word*). The Lexicon Principle is thus formulated as follows:

(36) **Lexicon Principle**

$$entry \rightarrow HS_1 \vee HS_2 \vee \ldots \vee HS_n$$

Objects of type *entry* introduce attributes PHON (with values of type *list*(*orth*)), CONT (with values of type *content*) and HEAD (with values of type *head*).

(37)
$$\begin{bmatrix} entry \\ \text{PHON } list(orth) \\ \text{HEAD } head \\ \text{CONT } content \end{bmatrix}$$

In the simplest cases, objects of type *word* may take their attribute values directly from the ENTRY structure which will be now a part of the *word* structure.

2.6 Modifiers

In HPSG, modifiers (adjuncts) are normally represented via the attribute MOD of type *head*, whose value is a (at most one element) list of objects of type *synsem*. We divide this *synsem* information into syntactic and semantic parts. Thus, the attribute MOD has values of type *mod*, which has two attributes: SYN of type *head* and SEM of type *content*:

(38)
$$\begin{bmatrix} mod \\ \text{SYN } head \\ \text{SEM } content \end{bmatrix}$$

(39)
$$\begin{bmatrix} head \\ \text{MOD } list(mod) \\ \text{ARG-ST } \begin{bmatrix} arg\text{-}st \\ \text{SUBJ } list(synsem) \\ \text{ARGS } list(synsem) \end{bmatrix} \end{bmatrix}$$

Since, in Polish, we do not observe any clear syntactic differences between complements and modifiers, we adopt here the solution known in HPSG as "adjuncts-as-complements" (see [BMS01], [Prz99]). The idea consists in placing modifiers together with arguments on the ARG-ST|ARGS list.

Taking the description of a word from a lexicon, beside taking the appropriate word's arguments form the ENTRY structure, we add to the ARG-ST|ARGS list a list of (4 in (41)). Moreover, according to the modifiers type the value of the attribute CONT can also be changed., (see "f(3,4)" in (41)).

Since the only part of the HEAD value of the ENTRY structure which can be changed is inside the ARG-ST attribute, we divide HEAD structure into ARG-ST attribute and MORSYN attribute containing all remaining head attributes:

(40)
$$\begin{bmatrix} head \\ \text{ARG-ST}\ \ arg\text{-}st \\ \text{MORSYN}\ \begin{bmatrix} morsyn \\ \text{MOD}\ \ list(mod) \end{bmatrix} \end{bmatrix}$$

Applying all the modifications just introduced, we can formulate the constraint describing the relation between ENTRY and SYNSEM|LOCAL structures in the following way:

(41)
$$word \rightarrow \begin{bmatrix} \text{PHON}\ \boxed{1} \\ \text{SS|LOC}\ \begin{bmatrix} \text{CAT|HEAD}\ \begin{bmatrix} head \\ \text{MORSYN}\ \boxed{6} \\ \text{ARG-ST}\ \begin{bmatrix} arg\text{-}st \\ \text{SUBJ}\ \boxed{5} \\ \text{ARGS}\ \boxed{2} \oplus \boxed{4} list(\begin{bmatrix} \text{MOD}\ \langle \begin{bmatrix} \text{SYN}\ \boxed{6} \end{bmatrix} \rangle \end{bmatrix}) \end{bmatrix} \end{bmatrix} \\ \text{CONT}\ f(\boxed{3},\boxed{4}) \end{bmatrix} \\ \text{ENTRY}\ \begin{bmatrix} entry \\ \text{PHON}\ \boxed{1} \\ \text{HEAD}\ \begin{bmatrix} head \\ \text{MORSYN}\ \boxed{6} \\ \text{ARG-ST}\ \begin{bmatrix} \text{SUBJ}\ \boxed{5} \\ \text{ARGS}\ \boxed{2} \end{bmatrix} \end{bmatrix} \\ \text{CONT}\ \boxed{3} \end{bmatrix} \end{bmatrix}$$

3 Selected Phenomena of Polish

3.1 Agreement

One of the main grammatical issue in Polish is agreement[4]. We concentrate on two main types of agreement: adjective–noun agreement and subject–verb agreement.

The adjective must agree with the noun in number, gender, and case, see (42). The same type of agreement takes place between the possessive pronoun and the noun (43), as well as between the numeral and the noun (44).

(42) a. pięknej dziewczynie
pretty$_{sg,fem,gen}$ girl$_{sg,fem,gen}$
b. *pięknemu dziewczynie
pretty$_{sg,masc,gen}$ girl$_{sg,fem,gen}$

(43) moją matką
my$_{sg,fem,inst}$ mother$_{sg,fem,inst}$

[4] The problem of agreement for Polish is discussed within the HPSG setup also in [Czu95] and [CP95].

(44) dwóch dziewczynach
two$_{fem,loc}$ girls$_{pl,fem,loc}$

The nominative subject agrees with the verb in person, number and gender, (45)

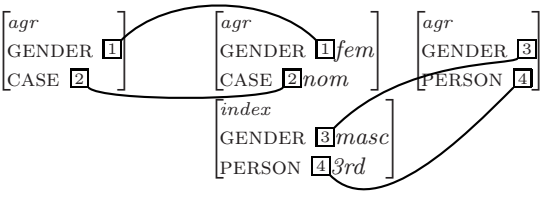
(45) a. Matka przyszła.
mother$_{sg,fem,nom}$ came$_{3rd,sg,fem}$
b. *Matka przyszedł.
mother$_{sg,fem,nom}$ came$_{3rd,sg,masc}$

The above examples show the common type of agreement in Polish. There are also untypical agreement where, e.g., semantic gender is not the same as the syntactic gender, and the verb agrees with the semantic gender of the noun, see (46).

(46) Jego wspaniałomyślna wysokość przyszedł.
His$_{fem,nom}$ magnanimous$_{sg,fem,nom}$ highness$_{sg,fem,nom}$ came$_{3rd,sg,masc}$
'His magnanimous highness came.'

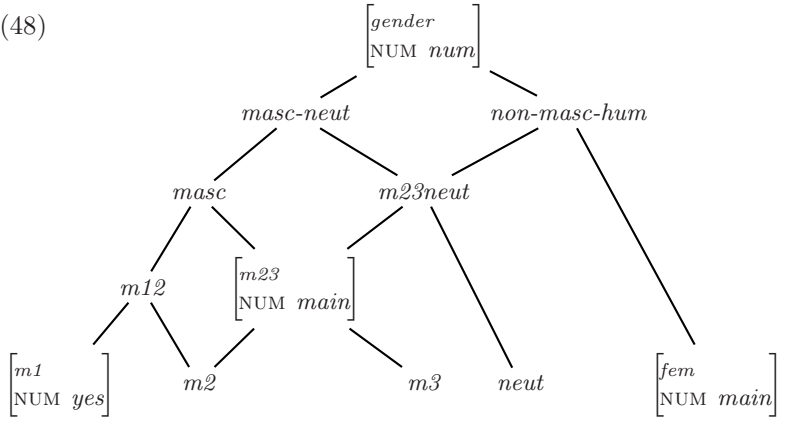
To cope with the problem of different syntactic and semantic gender of such nouns, the *index* structure (semantic gender, number and person) of the subject agrees with the syntactic gender, number and person (structure *agr*) of the verb, while the NP-internal agreement uses only syntactic attributes:

(47) Jego wspaniałomyślna wysokość przyszedł.



(48) presents the hierarchy of Polish gender elaborated to account for different types of agreement adjective–noun and numeral–noun agreement.

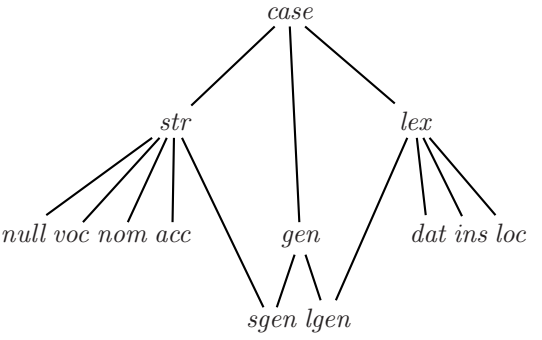
(48)



3.2 Case Assignment

The hierarchy of Polish cases[5] is given in (49). Cases are divided into structural and lexical types. The value of structural case is established not only by the subcategorisation rules but also by the environment, lexical cases are determined independently.

(49)



The most famous case phenomena in Polish is the Genitive of Negation (GoN), i.e., the shift of a direct object's case from accusative in a non-negated clause to genitive in the negated clause, see (50).

(50) a. Piszę listy.
write$_{1st,sg}$ letters$_{acc}$
'I am writing letters'
b. Nie piszę listów / *listy
NM write$_{1st,sg}$ letters$_{gen}$ / letters$_{acc}$
'I am not writing letters'

This phenomenon is nonlocal: in the case of the long distance Genitive of Negation, an argument of a *lower* verb may occur in the genitive when a *higher* verb is negated, see (51).

(51) Nie chciałem pisać listów / *listy.
NM want$_{1st,sg}$ write$_{inf}$ letters$_{gen}$ / letters$_{acc}$
'I didn't want write letters'

Interesting case assignment phenomena also include complex case patterns in numeral phrases, see (52)

(52) a. Pięć kobiet przyszło.
five$_{nom?/acc?}$ women$_{gen,pl}$ came$_{3rd,sg,neut}$
'Five women came.'
b. Rozmawiam z pięcioma kobietami / *kobiet.
talk$_{1st,sg}$ with five$_{ins}$ women$_{ins/*gen}$
'I am talking with five women.'

[5] The problem of case assignment is widely discussed in [Prz99].

Case patterns in predicative constructions in Polish are also interesting. In simple cases, predicative adjectives agree with predicated elements, see (53). But the predicative adjective can sometimes occur in the instrumental case, see (54).

(53) On jest miły.
he$_{nom}$ is nice$_{nom}$
'He is nice.'

(54) Pamiętam go miłego / miłym.
remember$_{1st,sg}$ him$_{acc}$ nice$_{acc}$ / nice$_{ins}$
'I remember him as nice.'

3.3 Binding Theory

The next problem addressed in the grammar is the binding theory for Polish[6]. It is not the whole theory but only Principles A and B formulated for pronominals and reflexive anaphors (possessive and non-possessive).

Anaphor binding in Polish can be roughly characterised as subject oriented and clause-bound. The distribution of personal pronouns in these sentences is complementary to that of anaphors, i.e., pronouns have to be disjoint with the subject, while coindexation with another non-subject argument of a verb or a clause external NP is correct, see (55)

(55) a. Jan$_i$ opowiadał Piotrowi$_j$ o sobie$_{i/*j}$/nim$_{*i/j}$.
John told Peter about self/him
'John told Peter about himself/him.'

b. Jan$_i$ powiedział, żeby Piotr$_j$ opowiedział o sobie$_{*i/j}$/nim$_{*i/j}$.
John told COMP Peter told about self/him
'John told Peter to tell about himself/him.'

The theory accounts for such important phenomena as medium distance binding in the case of control verbs, see (56). The possessive anaphor *swoje* has two possible antecedents: the sentential subject, *Jan*, or clause-internal one, *Piotrowi*. On the other hand, the possessive pronoun *jego* may not be bound by any of these elements.

(56) Jan$_i$ kazał Piotrowi$_j$ przynieść swoje$_{i/j}$/ jego$_{*i/*j}$ dokumenty.
John ordered Peter bring$_{inf}$ self's his documents
'John ordered Peter to bring his documents .'

We also analyse binding within noun phrases that can have subject (57) and attributive adjective phrases (58).

(57) wiara Marii$_i$ w siebie$_i$/ nią$_{*i}$
faith Mary's in self her
'Mary's faith in her (ability)'

[6] An HPSG theory of binding in Polish is presented in [Mar99,Mar01].

(58) Jan$_i$ zatelefonował do Piotra$_j$ napadniętego w swoim$_{*i/j}$/ jego$_{i/*j}$ domu.
John phoned to Peter robbed in self's his house
'John phoned Peter robbed in his house.'

A virtual subject is necessary to interpret the differences of binding in impersonal constructions, see (59).

(59) a. Kupiono sobie/ im lekarstwa.
bought self/ them medicines
'They bought medicines for themselves/ them.'

b.
$$\begin{bmatrix} \textit{word} \\ \text{PHON } \langle \text{kupiono} \rangle \\ \text{SS|LOC} \begin{bmatrix} \textit{local} \\ \text{CAT|HEAD|ARG-ST} \begin{bmatrix} \textit{arg-st} \\ \text{SUBJ } \langle \textit{pro}_j \rangle \\ \text{ARGS } \langle \text{NP}_k, \text{ NP: } \textit{pron}_i \rangle \end{bmatrix} \\ \text{CONT} \begin{bmatrix} \textit{buying} \\ \text{BUYER } j \\ \text{BOUGHT } k \\ \text{BENEFICIARY } i \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Theory of binding is defined on the ARG-ST structure. The most important relation, corresponding to local o-command relation (for English) [PS94, ch.6], is the relation of *local subject-command*, henceforth *local s-command* (see definition (61)). To formulate this relation, it is convenient to introduce a class of transparent phrases whose boundaries can be crossed in the process of binding.

(60) A *synsem* object X is *transparent* if X is a PP, VP[inf] or an NP without subject.

The definition of local s-command is given in (61):

(61) Let Y and Z be *synsem* objects. Then Y *locally s-commands* Z in case either:
i. exists a *arg-st* structure for which Y belongs to its SUBJ, and Z belongs to the list of its ARGS; or
ii. Y locally s-commands a transparent X and Z belongs to the ARG-ST structure of X

The local o-binding and local o-freeness relations (for English) are substituted by local s-binding and local s-freeness, respectively, see definition (62).

(62) Y *locally s-binds* Z just in case Y and Z are coindexed and Y locally s-commands Z. If Z is not locally s-bound, then it is said to be *locally s-free*.

The Principles A and B for Polish are formulated as in (63)

(63) Principle A. A reflexive anaphor must be locally s-bound.
Principle B. A pronoun must be locally s-free with the exception of possessive pronouns in first and second person and when possessive pronoun is bound by explicit subject of NP.

3.4 Relative Clauses

Polish relative clauses can be illustrated by the examples given in (64).

(64)
a. ten$_{nom,sg,masc}$, komu$_{dat,sg,masc}$ zazdrościcie$_{pl}$ _$_{dat}$
one whom you$_{pl}$ envy
'someone you envy'

b. chłopak$_{nom,sg,masc}$, [którego$_{acc,sg,masc}$ siostrze]$_{dat,sg,fem}$ zazdrościcie _$_{dat}$
a boy whose sister you$_{pl}$ envy _
'a boy whose sister I envy'

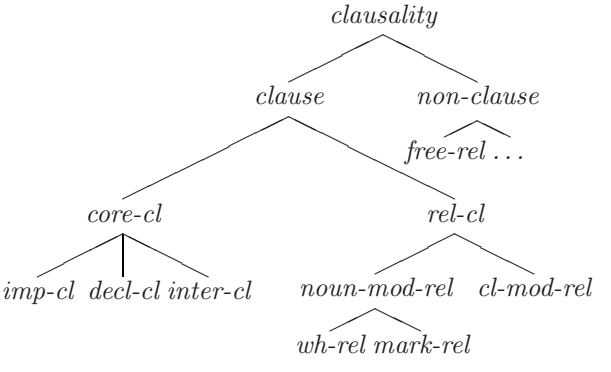
c. pióro$_{sg,neut}$, co nim$_{instr,sg,neut}$ /*() pisałam _$_{instr}$
a pen that with what /() I wrote
'a pen I wrote with'

d. Anna tańczyła, czemu Paweł przyglądał się uważnie.
Ann danced what$_{dat}$ Paul looked at carefully
'Paul looked carefully at Ann dancing'

e. ten/() kto sieje wiatr, zbiera burzę
this/() who$_{nom}$ sows a wind he picks a storm
'he who sows the wind shall reap the whirlwind'

The analysis of relative clauses presented here is based on the approach of [Sag97], which relies on multiple inheritance of constraints imposed on elements of phrase types hierarchy. However, we assume here only the phrase type hierarchy given in (28) while the clausality hierarchy is replaced by the CLAUSALITY attribute introduced for *phrases*. The possible values of this attribute are the subtypes of type *clausality* and are given in (65).

(65)



Clauses and non-clauses are distinguished on the basis of the type of the head element – for type *clause*, the value of the MORSYN attribute should be *personal, -no/-to, się, infinitival* or *marker*. Relative clauses are divided into *free-relatives*, which are a subtype of non-clauses, and relative clauses (proper), which are a subtype of clauses. The *clausality* hierarchy distinguishes relative clauses from core clauses on the basis of the *mod* attribute which is not empty for *rel-cl*. Relative clauses are then divided into those modifying noun phrases and those modifying clauses. Finally, noun modifying relatives are divided into

those starting with relative pronouns, those starting with the relative marker and the reduced relatives.

To account for non-local dependencies, we accept and extend the idea presented in [BMS01]. Information which is to be used non-locally is, as usual, grouped within the SYNSEM|NONLOCAL structure, (66). Here, apart from the SLASH and the REL attributes we define the RES attribute whose value is introduced by resumptive pronouns[7].

$$(66)\quad \begin{bmatrix} \textit{synsem} & \\ \text{LOCAL } \textit{local} & \\ \text{NONLOCAL} & \begin{bmatrix} \text{SLASH } \textit{list}(\textit{local}) \\ \text{REL } \boxed{1}\ \textit{list}(\textit{index}) \\ \text{RES } \boxed{2}\ \textit{list}(\textit{index}) \end{bmatrix} \end{bmatrix} \wedge \texttt{max-one}(\boxed{1}, \boxed{2}).^{8}$$

Phrases of the type *val-ph* inherit their NONLOCAL value from their head-daughters, while *fill-ph* phrases are places for binding nonlocal dependencies.

For a word, its SLASH value is obtained by gathering the SLASH values from all its dependents (if it has any), while the RES and REL values, apart from being gathered from all word's dependents, can also be specified within a lexicon description (inside the ENTRY structure). The way of computing the appropriate values are encoded in the Nonlocal Lexical Amalgamation Principle.

Polish relative pronouns can be divided into nominal relative pronouns: *który 'who/what', jaki 'which', kto 'who'* and *co 'what'* and adverbial relative pronouns, e.g., *gdzie 'when', kiedy 'where', skąd 'where from'*. In general, the use of the Polish relative pronouns is quite similar to other languages (e.g., Bulgarian, English). However, there are some specific features which have to be noted. One such idiosyncrasy is the use of different nominal relative pronouns (*który* vs. *kto*) in relation to different types of nominal phrases. Clauses beginning with *który* can modify noun phrases headed by common nouns, proper nouns, personal and demonstrative pronouns, (67a), while the relative pronouns *kto 'who'* and *co 'what'* can modify indefinite and negative pronouns, (67b).

(67) a. pies / Jan / on / tamten który / *kto biegnie
a dog / John / he / that KTóRY / *who runs

b. coś/nic czemu/*któremu się przyglądasz
something/nothing what/*KTóRY self you look at
something/nothing you look at

To give a complete analysis of Polish relative clauses beginning with pronouns, one has to deal with the following problems: ensuring the gender and number agreement between the modified noun and the pronoun, assigning the

[7] To make the formalization easier, we use lists instead of sets of values. In case of REL and RES this change is purely theoretical, as for Polish these attributes can have at most one element set (or list) as their value.

[8] Relation `max-one` represents the fact that, in Polish clauses, only one relative word or one resumptive pronoun can occur, so lists REL and RES can have in sum only one element.

correct case value to the nominal pronoun, ensuring that relative pronouns occur in he appropriate context. All these relations are represented by the appropratte constrains on the *wh-rel* type and lexical entries of relative pronouns.

The second type of noun modifying relative clauses are those begining with the relative marker *co*, (64c). In this clauses the modified object is repeated by a resumptive pronoun (unless it fulfils the role of a subject). Resumptive pronouns are all personal pronouns except their nominative and strong forms (if they exist). They have have two alternative lexicon entries – one with the empty RES value and second with one element on the RES list identified with the INDEX value. In subject *co*-relatives there is no resumptive pronoun in a subject position. To account for this fact, we assume that *pro* objects, which represent dummy subjects can be also interpreted as resumptive pronouns.

Since relative clauses beginning with the marker *co* and the relative pronoun *który* modify different noun phrases that these beginning with the relative pronouns *kto* and *co* we divide *index* into two subtypes: *inst-index*, which will be assigned to all nouns which can be modified by *który*, and *noninst-index* which is appropriate for indefinite and negative pronouns and relative pronouns *kto* and *co*.

As we analyse relative caluses as modifiers, we have to accept non-empty MOD value of verbal phrases. However, we would like to exclude sitations in which "an ordinary " clause (e.g., *Jan śpi 'John sleeps'*) is a modifier. Our solutions consists in changing the scope of the Head Feature Principle for the *fill-ph* phrases to all HEAD attributes besides MOD.

All constraints imposed on the types representing Polish relative clauses are given in [PKMM01], some previous work on the subject can be found in [Myk00].

3.5 Negation

There are several problems which are connected with the issue of negation in Polish[9]. Several words called *n-words* can appear only in the sentence where the environment is negated, it means that the verb is negated (68), or adjective has negative meaning (69) or that there is a negative preposition in the sentence (70). The examples of n-words are *nikt* nobody, *nigdy* never, *żaden* (none).

(68) Nikt *(nie) dał Marysi książki.
nobody NM gave Mary book
'Nobody gave Mary a book.'

(69) Westchnął *(nie)zauważalnie dla nikogo.
sight NM-noticably for nobody
'He sight without being noticed by anybody.'

(70) Zaczął bez żadnych wstępów.
started without none introductions
'He started straight away.'

[9] There are following papers connected with this subject: [PK97b,PK97a,PK99].

We can say that n-words in Polish are sensitive for negation. This feature is represented by attribute NEG-SENS which indicates that a word needs negative environment or does not need. In (71) is given the description of the n-word *nikt* (nobody).

(71)
$$\begin{bmatrix} \textit{entry} \\ \text{PHON } \langle \textit{nikt} \rangle \\ \text{HEAD} \begin{bmatrix} \text{MORSYN} \begin{bmatrix} \textit{n-noun} \\ \text{CASE } \textit{nom} \end{bmatrix} \\ \text{ARG-ST} \begin{bmatrix} \text{SUBJ } \langle\rangle \\ \text{COMPS } \langle\rangle \end{bmatrix} \end{bmatrix} \\ \text{CONT} \begin{bmatrix} \textit{ppro} \\ \text{INDEX} \begin{bmatrix} \text{PER } \textit{3rd} \\ \text{NUM } \textit{sg} \\ \text{GEND } \textit{m1} \end{bmatrix} \\ \text{RESTR } \{\} \end{bmatrix} \\ \text{NEG-SENS } + \end{bmatrix}$$

To represent negative environment we use the attribute POLARITY. For example in (72) there is represented semantics of the following events: *lubi* (like) and *nie lubi* (does not like).

(72) a. lubi: $\begin{bmatrix} \textit{psoa} \\ \text{NUCL } \textit{like} \\ \text{POLARITY } + \end{bmatrix}$

b. nie lubi: $\begin{bmatrix} \textit{psoa} \\ \text{NUCL } \textit{like} \\ \text{POLARITY } - \end{bmatrix}$

The negative concord principle is defined in the book in order to attain the correspondance between an n-words and its environment.

3.6 Coordination

Coordinated structures are widespread in natural languages but their formal analysis presents many problems. It is possible to coordinate not only phrases of the same categories, but also different categories can be conjoined as well. Although it is often assumed that coordination may apply only to constituents, coordination of non-constituents or partial constituents (phrases which share arguments) is not uncommon in natural languages. In the book[10], we restrict ourselves only to constituent coordination. For this reason, we base our analysis on the HPSG account of constituent coordination presented in [Par92]. The analysis captures coordination of partial (unsaturated) constituents as well, see (73).

(73) Jan przeczytał i zrecenzował artykuł.
John read and reviewed paper

[10] This problem is also discussed in [KMMP00,KMM00].

We deal with coordination of unlike categories but only in the case of (verbal) modifiers, see (74), where the adverb *szybko* is coordinated with the preposition phrase *bez zastanowienia*.

(74) Odpowiadał szybko i bez zastanowienia.
answered-he quickly and without thinking

Following [Par92], we treat conjunction as a functional head of the coordinated phrase and coordinated elements as complements. In the book there are discussed several types of conjunctions: 'monosegmental' conjunctions, e.g., *i* 'and', *lub* 'or', etc, we consider also discontinuous conjunctions, such as *zarówno . . . jak też* 'both . . . and', *nie tylko. . . lecz również* 'not only. . . but also'.

Let us see the lexical entry for *zarówno. . . jak też* 'both. . . and', (75).

(75)
$$\begin{bmatrix} \textit{entry} \\ \text{PHON } \langle \textit{zarówno, jak i} \rangle \\ \text{HEAD}|\text{CONJ } \textit{conj} \\ \text{NEG-SENS } - \end{bmatrix}$$

In the (76) there is the example of anaysies of the phrase *zarówno Ania jak też Adam* 'both Ania and Adam'

(76)
$$\begin{bmatrix} \textit{phrase} \\ \text{PHON } \langle \textit{zarówno, Ania, jak i, Adam} \rangle \end{bmatrix}$$

$$\begin{bmatrix} \text{PHON } \langle \textit{Ania} \rangle \\ \text{SS } \boxed{1}\textit{canon-ss} \end{bmatrix} \quad \begin{bmatrix} \textit{word} \\ \text{PHON } \langle \textit{zarówno, jak i} \rangle \\ \text{ARG-ST}|\text{ARGS } \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \quad \begin{bmatrix} \text{PHON } \langle \textit{Adam} \rangle \\ \text{SS } \boxed{2}\textit{canon-ss} \end{bmatrix}$$

The correct order of conjuncts and the conjunction is obtained via a general constrained which adds the first phonological 'segment' of the conjunction to the phonology of the first conjunct while the second phonological 'segment' of the conjunction precedes the second conjunct.

In languages with a rich morphological system agreement patterns in coordinated structures is quite complex. In the book, we present some aspects of agreement in Polish coordinated NPs and what kind of restriction are undertaken in the grammar. We assume that coordinated phrase must agree with verbs in plural number, see (77), so we are not able to analyse correct sentence (78).

(77) Przyszli Jan i Maria.
came$_{pl,3rd}$ John and Mary
'John and Mary came.'

(78) Przyszedł Jan i Maria.
came$_{sg,masc,3rd}$ John and Mary
'John and Mary came.'

There are also defined relations *gender* (79) and *min* which assign respectively the proper gender and person of the verb connected with coordinated phrase.

(79) `gender(`*masc-hum*`,` *gender*`,` *masc-hum*`).`
`gender(`*gender*`,` *masc-hum*`,` *masc-hum*`).`
`gender(`*non-masc-hum*`,` *non-masc-hum*`,` *non-masc-hum*`).`

This relation are necessary to analyse following sentences:

(80) Chłopiec i dziewczynka biegali po parku.
boy_{masc} and girl_{fem} $\text{run}_{masc-hum}$ in park
'A boy and a girl were running in the park.'

(81) Ja i ty przyszliśmy / *przyszliście.
I_{1st} and you_{2nd} came_{1st}-we came_{2nd}-you
'I and you came.'

The approach presented in the book captures only several types of coordinated phrases but it allows us to apply general HPSG grammatical principles both to coordination and to other types of phrases.

4 Conclusion

In this paper, we have presented a range of phenomena typical for Polish and indicated ways of analysing those phenomena within HPSG. We have also presented modifications of the standard [PS87,PS94] HPSG theory useful or necessary to formulate a straightforward account of Polish syntax. The diversity of the phenomena accounted for, involving both textually frequent phenomena (negation, relative clauses, simple case assignment) and textualy untypical phenomena (idiosyncratic patterns of case assignment, special cases of binding and agreement), lead to the conclusion that Head-driven Phrase Structure Grammar is a formalism well-suited for analysing morphologically-rich "free word-order" languages such as Polish.

The account alluded to above, fully described in [PKMM01] and references therein, has been partially implemented in ALE [CP01]. The implementation varies from the theoretical analysis in many respects due to the underlying differences between ALE and HPSG. In particular, two versions of the grammar have been implemented: a version close to the linguistic theory which, however, led to a much less efficient implementation, and a more efficient version taking into account various non-HPSG mechanisms offered by ALE. We intend to extend the work reported here by constructing an HPSG-based parser of Polish going well beyond the empirical boundaries of the current HPSG grammar presented above.

References

[BMS01] Gosse Bouma, Robert Malouf, and Ivan A. Sag. Satisfying constraints on extraction and adjunction. *Natural Language and Linguistic Theory*, 19(1):1–65, 2001.

[BP99] Robert D. Borsley and Adam Przepiórkowski, editors. *Slavic in Head-Driven Phrase Structure Grammar*. CSLI Publications, Stanford, CA, 1999.

[BP00] Piotr Bański and Adam Przepiórkowski, editors. *Proceedings of the First Generative Linguistics in Poland Conference*, Warsaw, 2000. Institute of Computer Science, Polish Academy of Sciences.

[CP95] Krzysztof Czuba and Adam Przepiórkowski. Agreement and case assignment in Polish: An attempt at a unified account. Technical Report 783, Institute of Computer Science, Polish Academy of Sciences, 1995.

[CP01] Bob Carpenter and Gerald Penn. *The Attribute Logic Engine (Version 3.2.1). User's Guide*. Carnegie Mellon University, Pittsburgh, December 2001.

[Czu95] Krzysztof Czuba. Zastosowanie dziedziczenia do analizy wybranych aspektów języka polskiego. Master's thesis, Uniwersytet Warszawski, Warsaw, 1995.

[KMM00] Anna Kupść, Małgorzata Marciniak, and Agnieszka Mykowiecka. Constituent coordination in Polish: An attempt at an HPSG account. In Bański and Przepiórkowski [BP00], pages 104–115.

[KMMP00] Anna Kupść, Małgorzata Marciniak, Agnieszka Mykowiecka, and Adam Przepiórkowski. Składniowe konstrukcje współrzędne w języku polskim: Próba opisu w HPSG. Technical Report 914, Institute of Computer Science, Polish Academy of Sciences, 2000.

[Mar99] Małgorzata Marciniak. Toward a binding theory for Polish. In Borsley and Przepiórkowski [BP99], pages 125–147.

[Mar01] Małgorzata Marciniak. *Algorytmy implementacyjne syntaktycznych reguł koreferencji zaimków dla języka polskiego w terminach HPSG*. Ph. D. dissertation, Polish Academy of Sciences, 2001.

[MS97] Philip H. Miller and Ivan A. Sag. French clitic movement without clitics or movement. *Natural Language and Linguistic Theory*, 15:573–639, 1997.

[Myk00] Agnieszka Mykowiecka. Polish relative pronouns: An HPSG approach. In Bański and Przepiórkowski [BP00], pages 124–134.

[Par92] Maike Paritong. Constituent coordination in HPSG. Technical Report CLAUS 24, Universität des Saarlandes, Saarbrücken, 1992.

[PK97a] Adam Przepiórkowski and Anna Kupść. Negative concord in Polish. Technical Report 828, Institute of Computer Science, Polish Academy of Sciences, 1997.

[PK97b] Adam Przepiórkowski and Anna Kupść. Verbal negation and complex predicate formation in Polish. In Ralph C. Blight and Michelle J. Moosally, editors, *Proceedings of the 1997 Texas Linguistics Society Conference on the Syntax and Semantics of Predication*, volume 38 of *Texas Linguistic Forum*, pages 247–261, Austin, TX, 1997.

[PK99] Adam Przepiórkowski and Anna Kupść. Eventuality negation and negative concord in Polish and Italian. In Borsley and Przepiórkowski [BP99], pages 211–246.

[PKMM01] Adam Przepiórkowski, Anna Kupść, Małgorzata Marciniak, and Agnieszka Mykowiecka. *Formalny opis języka polskiego: Teoria i implementacja*. Akademicka Oficyna Wydawnicza, 2001. In progress.

[Prz99] Adam Przepiórkowski. *Case Assignment and the Complement-Adjunct Dichotomy: A Non-Configurational Constraint-Based Approach*. Ph. D. dissertation, Universität Tübingen, Germany, 1999.

[Prz00] Adam Przepiórkowski. Long distance genitive of negation in Polish. To appear, *Journal of Slavic linguistics*, 2000.

[Prz01] Adam Przepiórkowski. ARG-ST on phrases headed by semantically vacuous words: Evidence from Polish. In Dan Flickinger and Andreas Kathol, editors, *Proceedings of the 7th International Conference on Head-Driven Phrase Structure Grammar*, pages 267–284. CSLI Publications, Stanford, CA, 2001.

[PS87] Carl Pollard and Ivan A. Sag. *Information-Based Syntax and Semantics, Volume 1: Fundamentals*. Number 13 in CSLI Lecture Notes. CSLI Publications, Stanford, CA, 1987.

[PS94] Carl Pollard and Ivan A. Sag. *Head-driven Phrase Structure Grammar*. Chicago University Press / CSLI Publications, Chicago, IL, 1994.

[Rea92] Mike Reape. *A Formal Theory of Word Order: A Case Study in West Germanic*. Ph. D. dissertation, University of Edinburgh, 1992.

[Sag97] Ivan A. Sag. English relative clause constructions. *Journal of Linguistics*, 33(2):431–483, 1997.

[Świ92] Marek Świdziński. *Gramatyka Formalna Języka Polskiego*, volume 349 of *Rozprawy Uniwersytetu Warszawskiego*. Wydawnictwa Uniwersytetu Warszawskiego, Warsaw, 1992.

[Szp86] Stanisław Szpakowicz. *Formalny opis składniowy zdań polskich*. Wydawnictwa Uniwersytetu Warszawskiego, Warsaw, 1986.