Agnieszka Mykowiecka, Małgorzata Marciniak, Adam Przepiórkowski, Anna Kupść,
Institute of Computer Science, PAS, Warsaw

An Implementation of a Generative Grammar of Polish

Introduction

In spite of the great amount of research being done in the field of computational linguistics, computational work on Polish syntax lags behind that on other European languages. There are only two documented formal grammars of Polish (Szpakowicz, 1978; Świdziński, 1992), both of which are based on the Metamorphosis Grammar formalism (i.e., to all intents and purposes, Definite Clause Grammar). In this paper we present an implementation of a grammar of Polish formalised within Head-driven Phrase Structure Grammar (HPSG), a modern generative linguistic formalism with sound logical foundations.

The grammar briefly described here stems from exhaustive analyses of such phenomena as case assignment, relative constructions, binding, negation, etc., by particular authors of this paper. These subtheories were subsequently combined into a coherent and relatively large grammar of Polish. The effect of this efforts is described in detail in (Przepiórkowski et al., 2002) and summarised in (Bolc, 2002). The present paper is a short presentation of the implementation based on that theory.

The choice of the HPSG formalism to describe Polish syntax was motivated by several promising features of the theory. The most important is the explicitness and formality that made HPSG one of the most popular linguistic formalisms in computational linguistic applications. HPSG is a comprehensive linguistic formalism for work on syntax, morphology and semantics, as well as phonology and pragmatics. It is a formalism, i.e., a set of formal tools for representing linguistic analyses and also a linguistic theory, i.e., a collection of analyses of various phenomena described using this formalism. The grammar presented here differs in many ways from the description of (Pollard and Sag, 1994) and these differences will be pointed out at various places within the presentation of the implementation. Additional modifications of the theory were necessary in order to implement the grammar with the use of ALE (Carpenter and Penn, 2001), an engine for implementing unification-based attribute-value grammars similar, but not exactly equivalent to HPSG grammars.

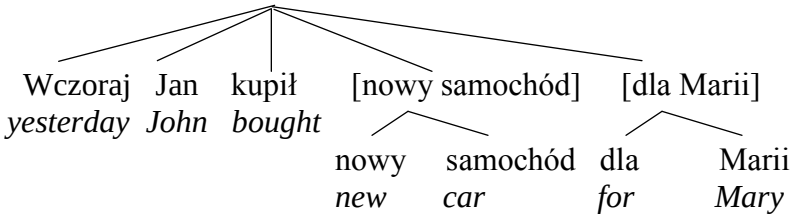
ALE — The Attribute Logic Engine — was created at the Carnegie Mellon University by Gerald Penn and Bob Carpenter. ALE is written in Prolog and can be used either with SICStus Prolog or with SWI Prolog. In our implementation we use SICStus Prolog, version 3.5.8. ALE does not supply any graphical interface, so a visualisation tool developed by Gertjan van Noord, HDRUG, was employed. Both systems, ALE and HDRUG, are publicly available from www pages.

Phrase structure

It is generally accepted, both in HPSG and in other generative formalisms, that a head element of a phrase (e.g., the verb *kupił* 'bought' within the phrase *Jan kupił Marii samochód* 'John bought

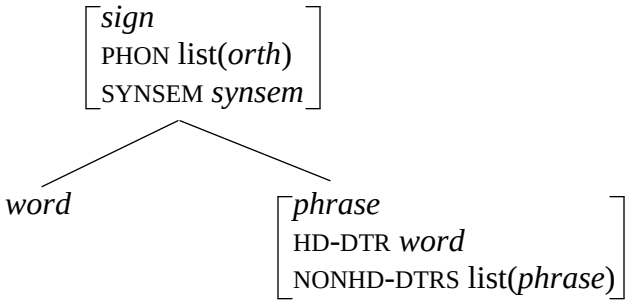
Mary a car') takes its complements (e.g., nouns *Marii* 'Mary' and *samochód* 'a car') first, forming an almost saturated phrase (e.g., the verb phrase *kupił samochód Marii* 'bought Mary a car'). Subsequently, this phrase combines with the subject (if any) and forms a saturated phrase (e.g., a clause *Jan kupił Marii samochód* 'John bought Mary a car'). Arguments for these two stages of phrase construction do not immediately carry over to Polish, where sentences in which the subject is realised 'closer' to a verb than its complements are well-formed (e.g. *Wczoraj kupił Jan nowy samochód dla Marii* 'Yesterday John bought a new car for Mary'.) For this reason, we assume that, in Polish, all arguments of a head are realised simultaneously. Moreover, and again contrary to tradition, we posit that modifiers are attached at the same level of tree structure as complements. As a result, phrase structures constructed by the grammar are relatively flat, as illustrated in the picture below:

(1) Wczoraj Jan kupił nowy samochód dla Marii



In order to ensure simultaneous realisation of all dependents, we require the head element of a phrase to be a *word* (not a *phrase*), while elements of the NONHD-DTRS list must be *phrase*s (not *word*s), see (2) below. As a consequence, there is no partially saturated phrase. Additionally, *word*s which do not subcategorise for any arguments can become syntactic arguments of other *word*s only if 'type-shifted' to *phrase*s via a unary syntactic rule. Since any linguistic object (that is a *sign*) must be either a head (that is a *word*) or a non-head element (that is a *phrase*), spurious ambiguity is avoided.

(2)



Type hierachies like (2) above define the space of possible linguistic structures: they specify possible types of linguistic objects (e.g., *phrase* or *word*) and attributes these objects may have. In HPSG *qua* linguistic formalism, as in all constraint-based formalisms, this space of potentially possible structures is constrained by principles (or constraints) of the grammar which eliminate ill-formed constructions. However, ALE, the tool employed for implementation in this project, differs from this purely constraint-based approach: for reasons of efficiency, principles of the HPSG grammar must be expressed as unification-based definite-clause rules. This difference is discussed in the next subsection.

Representation of dependent elements — VALENCE and ARG-ST

As standard in HPSG, we use two attributes VALENCE and ARG-ST to represent dependents of a linguistic object. However, these attributes are introduced and employed differently. Since all arguments are realised simultaneously, the standard split of VALENCE into SUBJ and COMPS lists is not necessary. Only *word*s may be combined with other elements and, hence, we assume that VALENCE is appropriate only for *word*s and its value is a list of *synsems*:

(3) $\begin{bmatrix} \textit{word} \\ \text{VALENCE} \;\; \text{list}(\textit{synsem}) \end{bmatrix}$

The above changes result in the following reformulation of the Valence Principle:

(4) $\textit{phrase} \rightarrow \begin{bmatrix} \text{HD-DTR} \mid \text{VALENCE} \; \boxed{1} \\ \text{NONHD-DTRS} \; \boxed{2} \end{bmatrix} \wedge \text{synsems-signs}(\boxed{1}, \boxed{2})$

where synsems-signs (1 , 2) is true when 2 is a list of *sign* objects with their SYNSEM values equal to subsequent elements of the 1 list.

The Valence Principle is implemented in ALE as a rule (schema1 rule below) with procedural attachments. As Polish word order is relatively free, the schema permits all orders of phrase elements, apart from those explicitly excluded. The *left_right_dtrs* clause requires arguments of a preposition and a marker to follow the head.

(5)
```
schema1 rule
 (Mother, val_ph,
  hd_dtr: HDtr,
  non_hd_dtrs: NHdtr)
  ==>
  cats> NHdtr1,
  cat> (HDtr, synsem:loc:cat:head:(Head, morsyn:M, val: Val)),
  goal> (argst_principle(Head,Val),
         word_principles_rest(HDtr),
         left_right_dtrs(M,NHdtr1,NHdtr2),
         synsems_to_signs(Val,Slist),
         permutation(Slist,NHdtr),
         append(NHdtr1,NHdtr2,NHdtr)),
  cats>  NHdtr2,
  goal>  (phrase_principles(Mother)).
```

The VALENCE attribute contains information about overtly realised arguments whereas information about other dependents, e.g., modifiers, is present on the ARG-ST list (see subsections below).

Traditionally such phenomena as agreement, case assignment or binding theory, are accounted for using the ARG-ST attribute defined only in *words* as a list of *synsems*. We also adopt this attribute but in a modified way. First, although the distinction between the subject and complements plays no role in describing the sentence structure, it is important for representing other phenomena, e.g., binding. Thus, the subject is distinguished in the ARG-ST, see (6). Second, we define ARG-ST as appropriate for the *head* objects. As a consequence, the ARG-ST attribute is

specified not only for *words* but also for *phrases* (see (Przepiórkowski, 2000) for a detailed discussion).

(6)
$$\begin{bmatrix} \textit{category} \\ \text{HEAD} \begin{bmatrix} \textit{head} \\ \text{ARG-ST} \begin{bmatrix} \textit{arg-st} \\ \text{SUBJ list}(\textit{synsem}) \\ \text{ARGS list}(\textit{synsem}) \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The attributes ARG-ST and VALENCE include similar but not necessarily the same information. The VALENCE list contains only those elements from the ARG-ST|SUBJ and ARG-ST|COMPS lists which are realised locally, while ARG-ST lists specify all predicate's dependents including those which are not syntactically realised. In order to handle non-overt arguments, we introduce (after Miller and Sag, 1997, Sag, 1997 and Bouma, Malouf and Sag, 2001) two subtypes of the *synsem* type: *canonical-synsem* (*canon-ss*), representing dependents which are locally realised and *noncanonical-synsem* (*noncanon-ss*), representing non-realised dependents. Now, the SYNSEM attribute takes objects of the *canon-ss* type as its value. This automatically makes all VALENCE elements have their SYNSEM values of the type *canon-ss* and *noncanon-ss* objects appear only on ARG-ST lists.

The Arguments Structure Principle representing the relation between ARG-ST and VALENCE attributes is given below: (⊕ is a shorthand for the append relation, while ○ is an infix notation of the shuffle relation (Reape, 1992) which inserts elements of one list between elements of another list):

(7) Arguments Structure Principle

$$\textit{word} \rightarrow \begin{bmatrix} \text{SYNSEM|LOCAL|CAT|HEAD|ARG-ST} & \begin{bmatrix} \textit{arg-st} \\ \text{SUBJ } \boxed{1} \\ \text{ARGS } \boxed{2} \end{bmatrix} \\ \text{VALENCE } \boxed{3} & \end{bmatrix}$$

$$\wedge\ \boxed{1} \oplus \boxed{2} = \boxed{3} \circ \text{list}(\textit{noncanon-ss})$$

This principle is implemented by the following ALE clause called in the schema1 rule.

(8)
```
argst_principle((argst:(subj:X,  args:Y)), Z)
                if append(X,Y,New), take_canons(New,Z).
take_canons([],[]) if true.
take_canons([noncanon_ss_t|R],Z) if take_canons(R,Z).
take_canons([(A,canon_ss_t)|R],[A|Z]) if take_canons(R,Z).
```

Non-canonical elements

There are three different kinds of phenomena which we describe using non-canonical arguments. The first one concerns the subject in two cases: the pro-dropped NP subject of personal verbs and the unrealised subject of non-personal verb forms. The unrealised subjects are represented by objects of the type *pro*, (9)

(9) a. Idzie do kina.
'(He) goes to the cinema'
b. Kazał Janowi wstać.

‘(He) told John to get up’

c. Ewa słuchała pijąc kawę.
‘Eve was sitting drinking coffee.’

The second kind of non-canonical objects, *gap*, are used to represent cases where arguments of verbs located ‘lower’ in the syntactic tree are realised ‘higher’ in the syntactic structure, e.g. (10).

(10) a. Kogo chciałeś, żebym zobaczył _?
Who (you) wanted that (me) see _?
‘Who did you want me to see?’

b. ... facet, o którym mówiłeś, że _ przyszedł za wcześnie.
... guy about who (you) said that _ came too early
‘a guy who you said that came too early’

The two kinds of *noncanonical synsem* types described above are typical for HPSG analyses. However, in our grammar the third kind of *noncanonical synsem* is introduced. We assume that some arguments may be ‘raised’ higher in the syntactic tree, and treated as if they were arguments of a higher verb. In the example below, arguments of the infinite *napisać* ‘*write*’ can be realised locally or they can be raised and analysed as arguments of the verb *planował* ‘*planned*’:

(11) Jan planował napisać książkę$_{acc}$.
‘John planned to write a book.’

A motivation for introducing argument raising in the case of infinitival complements comes form the genitive of negation phenomenon. If the higher verb (i.e., *planował*) is negated, the argument *książka* should occur in genitive rather than accusative (12a) just like in case when the lower verb *napisać* is negated, (12b):

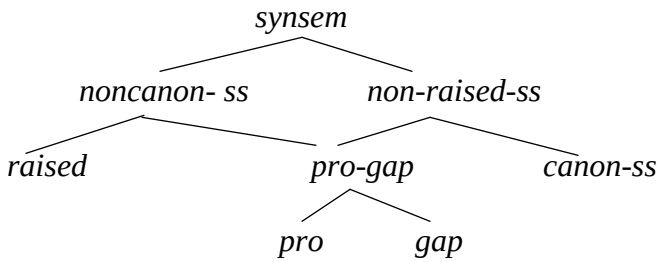
(12) a. Jan nie planował napisać książki$_{gen}$.
‘John did not plan to write a book.’

b. Jan napisał książkę$_{acc.}$ / Jan nie napisał książki$_{gen.}$
‘John wrote a book. / John did not write a book.’

We assume that case assignment principles are local, i.e., they operate on a single ARG-ST. To represent case changes when a higher verb is negated, we have to make arguments of the lower verb arguments of the higher verb. What is more, in some situations, the case change is optional, i.e., arguments of the lower verb may remain in accusative, (Przepiórkowski, 1999). To enable both analyses, we assume that argument raising is optional, i.e., examples like (12a) have several parses and differ in infinitival's arguments placement.

To represent raised objects, we introduce the *raised* vs. *non-raised* distinction. All *raised* objects are of *noncanon-ss* subtype while *non-raised* arguments are divided into *canon-ss*, *pro* and *gap* subtypes. The *synsem* hierarchy looks then as follows:

(13)



Raised arguments have to appear on the ARG-ST of a higher verb. The type of a raised argument is not passed higher up and, hence, only the local structure is shared between two ARG-STs. The ARG-ST of the verb *planował* is given below, (14). Arguments of the infinitival complement which are 'raised' are added (to be precise only their LOCAL values) to the arguments of the verb *planować.* The subject of an infinitive is never locally realised; the SUBJ value is a *noncanon-ss* object.

(14)

$$\left[\begin{array}{l} \textit{arg-st} \\ \text{SUBJ} \ \langle \left[\begin{array}{l} \textit{synsem} \\ \text{L|C|H|MS} \left[\begin{array}{l} \textit{noun} \\ \text{CASE } \textit{nom} \end{array}\right] \end{array}\right] \\ \text{ARGS} \ \langle \left[\begin{array}{l} \textit{synsem} \\ \text{L|C|H|MS} \left[\begin{array}{l} \textit{infinitive} \\ \text{ARG-ST} \left[\begin{array}{l} \textit{arg-st} \\ \text{SUBJ} \ \langle \textit{noncanon-ss} \rangle \\ \text{ARGS list}(\textit{non-raised-ss}) \circ \boxed{2} \text{list}(\textit{raised}) \end{array}\right] \end{array}\right] \end{array}\right] \oplus \boxed{2'} \end{array}\right]$$

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

Introducing raised arguments does not impose any changes on scheama1. However, to ensure that complete utterances do not have any raised elements left, we introduce the boolean ROOT attribute appropriate for *sign*s. Its value is '+' only if there are no raised elements and only such linguistic objects are considered complete utterances.

Two types of phrases

In our grammar, two phrase types are distinguished: *valence-phrase* and *fill-phrase* (named *val-ph* nad *fill-ph* respectively). Schema1 described above represents phrases which consist of the head element (a word) and its dependencies. This schema can be used to build not only clauses but also phrases with non-verbal head elements, e.g., noun or prepositional phrases. We employ it also for phrases with markers (complementizers) which are treated as heads. All phrases constructed according to this basic schema are called *val-ph* and the scope of the Valence Principle is limited to this type of phrases only.

The second phrase type — *filler-phrase*, as common in the HPSG tradition, is used to account for realisation of nonlocal dependents, for example:

(15) a. O kim chciałeś [żebym napisał _]?
'About whom did you wanted me to write?'
b. …facet, któremu mówiłeś, [że pożyczysz książkę _].

… a guy you said you would borrow the book to

In the above sentences, the elements *o kim* and *któremu* are non-locally realised complements of a verb. The implemented method of non-local dependencies analysis is based on (Bouma, Malouf and Sag, 2001). Complements that are not realised locally are realised 'virtually' by objects of the special type *gap* which have non-empty SLASH values. SLASH values are passed via the NONLOCAL attribute from word's dependents to the word itself and from the head element to the entire phrase. Filling of the gap needs matching the input element description with the value of the SLASH attribute. It is done by applying a special phrase schema given below:

(16)
```
schema11 rule
(Mother, fill_ph,
hd_dtr: HDtr,
non_hd_dtrs: [NHdtr],
synsem:nonloc: (slash:e_list,rel:e_list, res:e_list))
 ===>
cat> (NHdtr, synsem:loc:S,synsem:nonloc:slash:[]),
cat> (HDtr,   phrase_t,  synsem:nonloc:slash:[S|[]]),
goal>  phrase_principles(Mother).
```

Description of words

As the main goal of our project was to implement the grammar, we did not develop an elaborated lexicon. Only a few lexical entries were defined. They are included in the grammar code in the form defined by ALE. Lexical descriptions are of the type *entry_t* and are converted into *word_t* by the following <word> schema:

(17)
```
'<word>' rule
      (word_t, W,
       synsem:loc:(cat:head: (morsyn: X,
                              conj:Cn,
                              mod:Y,
                              argst: (subj: S, args: AS)),

                   cont: C),
       entry: HDtr)
       ===>
       cat> (HDtr,
             entry_t,
           lex_head: (morsyn: X,
                      conj:Cn,
                      mod:Y,
                      argst: (subj: S,  args: A)),
           lex_cont: C),
       goal> (canon(M),
              add_modifier(X,M),
              append(A,M,AS),
              append(AS,S,ASS),
              ini_nl(W,ASS)).

       ini_nl((synsem:nonloc:(slash:e_list,
```

```
                        rel:X,
                        res:RS),
          entry:(lex_rel:R,
                 lex_res:RS)),[]) if true.
ini_nl(_, [Y|Z]) if true.
add_modifier(_,[]) if true.
add_modifier(X,[(loc:(cat:head:(mod:[syn:X])))]) if true.
```

The schema above incorporates the idea of treating "adjuncts-as-complements" (see Bouma, Malouf and Sag, 2001 and Przepiórkowski, 1999). This idea consists in placing modifiers together with arguments on the ARG-ST|ARGS list and it is motivated by the fact that in Polish, we do not observe any syntactic differences between complements and modifiers. In the ALE implementation, for efficiency reasons, we had to limit the number of added modifiers to just one.

Agreement

Agreement is a very important grammatical issue in Polish[1]. Our theory describes agreement between adjectives and noun phrases as well as between subjects and verbs. An adjective phrase must agree with the noun phrase it modifies in number, gender, and case, see (18).

(18) a. starej kobiecie
old $_{sg,fem,gen}$ women $_{sg,fem,gen}$
b. *staremu kobiecie
old $_{sg,fem,gen}$ women $_{sg,fem,gen}$

The same type of agreement takes place between a possessive pronoun and a noun, as well as between a numeral and a noun, see (19).

(19) a. twoją książkę
your $_{sg,fem,acc}$ book $_{sg,fem,acc}$
b. dwaj chłopcy
two $_{pl,m1,loc}$ boys $_{pl,m1,loc}$

The nominative subject agrees with a verb in person, number and gender, see (20).

(20) a. Ojciec przyszedł.
'Father $_{sg,m1,nom}$ came $_{3rd,sg,masc}$'
b. *Ojciec przyszła.
'Father $_{sg,m1,nom}$ came $_{3rd,sg,fem}$'

All above common types of agreements are implemented in our grammar. In the theory, we described also atypical agreement that occurs for phrases with different semantic and syntactic genders but it has not been implemented yet.

Case Assignment

Case assignment is carefully discussed in Przepiórkowski (1999). Polish morphological cases are divided into structural and lexical types. Lexical cases are assigned in a subcategorization frame

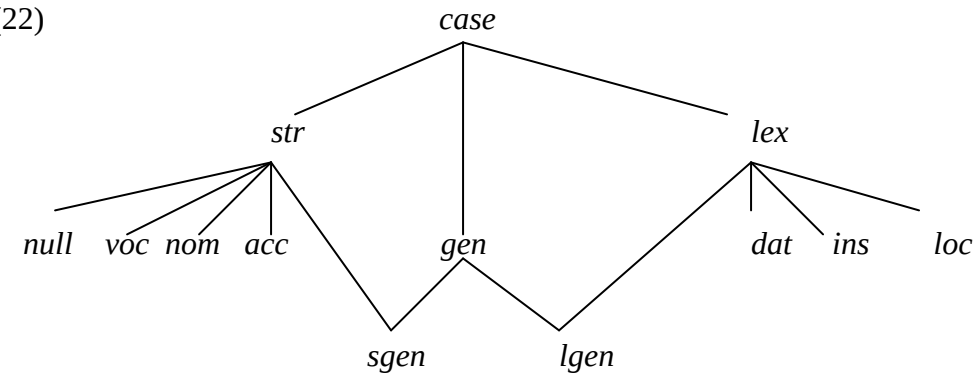
[1] Agreement for Polish within the HPSG is discussed in Czuba (1995) and Czuba and Przepiórkowski (1995).

of a predicate. Structural cases can be realised by different case forms depending on the context in which they occur. For example, in the sentences in (21) the complement's case changes according to presence/absence of negation on the verb *kupił*.

(21) a. On kupił kwiaty$_{acc}$.
'He bought flowers.'
b. On nie kupił kwiatów$_{gen}$.
'He didn't buy flowers.'

The hierarchy of Polish cases used in our grammar is given in (22).

(22)



The principles of case assignment for arguments of verbs, prepositions and nouns are implemented, but for predicative case assignment only a theoretical approach has been developed.

Binding Theory

In the theory, we formulated Principles A and B for binding of Polish[2] pronominals and anaphors, both possessive and non-possessive. Anaphora binding in Polish is in general subject-oriented and clause-bound. We analyse binding within several types of phrases, e.g., in personal verb phrases, (23a), impersonal verb phrases, see (23b). We described such important phenomena as medium distance binding in the case of control verbs, see (23c), and binding within noun phrases.

(23) a. Jan $_i$ spytał Piotra $_k$ o siebie $_{i / *k}$/niego $_{*i / k}$.
John asked Peter about self/him
'John asked Peter about himself/him.'
b. Oszukano siebie / ich..
cheated self / them
'They cheated themselves / them.'
c. Jan $_i$ kazał Piotrowi $_k$ kupić siebie $_{i / k}$ / jemu $_{*i / * k}$ nowy rower.
John ordered Peter buy self / him a new bicycle
'John ordered Peter to buy himself / him a new bicycle.'

The binding theory for Polish is inspired by an HPSG binding theory formulated for English in Pollard and Sag (1994, ch.6). The crucial relation for Polish is the local subject-command relation, see (25), corresponding roughly to the local o-command relation for English. We introduced a class of transparent phrases whose boundaries can be crossed in binding, (24).

[2] An HPSG binding theory for Polish is discussed in Marciniak (1999,2001).

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

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

The subject-command relation is defined recursively. For efficiency reasons, only one level of recursion has been implemented. Therefore pronominals in sentences (23 a-b) are correctly bound but binding in (26) is not captured in our implementation.

(26) Jan $_{i}$ spytał Piotra $_{k}$ o swoją $_{i / *k}$/jego $_{*i / k}$ siostrę.
John asked Peter about self's/his sister
John asked Peter about his sister.

Negation

One of the most important issues of Polish negation[3] are *n-words*, i.e., words such as *nikt* (*nobody*), *nic* (*nothing*), *żaden* (*none*). In Polish, *n-words* are licensed in a negative context, usually created by verbal negation, see (27), or the preposition *bez* (*without*), see (28).

(27) Nikt nie pytał o Piotra.
nobody NM asked about Peter
'Nobody asked about Peter.'

(28) Wstał bez niczyjej pomocy.
got up without nobodys help.
'He got up without nobodys help.'

Words' sensitivity to negation is indicated by the value of the NEG-SENS attribute. For *n-word*s the attribute is set to `+' which means that a word has to appear in a negative context, see description of the word *nic* in (29); NEG-SENS is `-' otherwise.

(29)
$$\begin{bmatrix} \textit{entry} \\ \text{PHON} \quad \langle \text{nic} \rangle \\ \text{HEAD} \begin{bmatrix} \text{MORSYN} \begin{bmatrix} \textit{n-noun} \\ \text{CASE} \ \textit{nom} \end{bmatrix} \\ \text{ARG-ST} \begin{bmatrix} \text{SUBJ} \ \langle\rangle \\ \text{COMPS} \ \langle\rangle \end{bmatrix} \end{bmatrix} \\ \text{CONT} \begin{bmatrix} \textit{ppro} \\ \text{INDEX} \begin{bmatrix} \text{PER} \ \textit{3rd} \\ \text{NUM} \ \textit{sg} \\ \text{GEND} \ \textit{neut} \end{bmatrix} \\ \text{RESTR} \ \{\} \end{bmatrix} \\ \text{NEG-SENS} \ + \end{bmatrix}$$

[3] For a more detailed discussion see, among others, Przepiórkowski and Kupść (1997, 1999).

Relative clauses

The analysis of relative clauses in Polish[4] is based on the approach presented in Sag (1997) which relies on multiple inheritance of constraints imposed on elements of phrase type hierarchy. Information about non-local dependencies is grouped within the NONLOCAL structure:

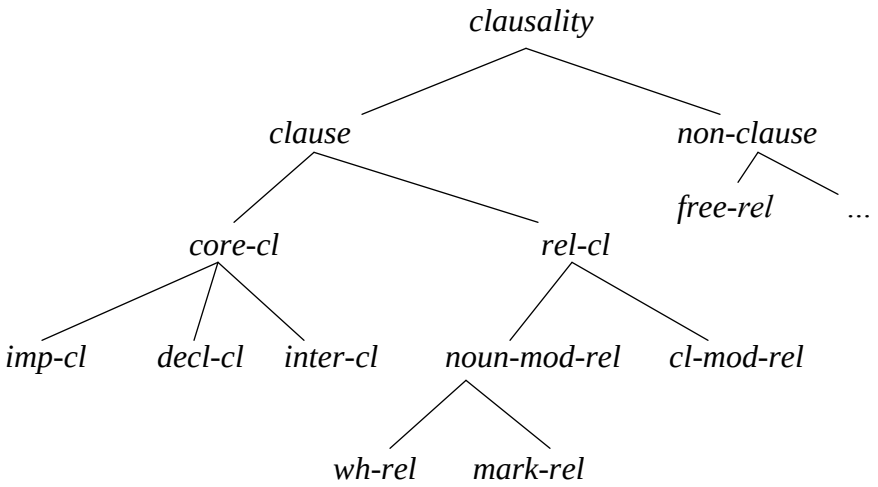
(30)
$$\begin{bmatrix} \textit{synsem} \\ \text{LOCAL} \quad \textit{local} \\ \text{NONLOCAL} \begin{bmatrix} \text{SLASH } \textit{list (local)} \\ \text{REL } \boxed{1} \quad \textit{list (index)} \\ \text{RES } \boxed{2} \quad \textit{list (index)} \end{bmatrix} \end{bmatrix} \quad \wedge \text{ max-one } (\boxed{1} , \boxed{2}).$$[5]

Relative pronouns introduce non-empty values of the REL attribute, while the RES attribute is defined specially for representing resumptive pronouns in sentences like (31).

(31) ten, co go zgubiłeś
one that it_{res} (you) lost
'the one you lost'

Relative clauses are divided into several subtypes forming the hierarchy given in (32). In the implemented grammar, the clause type is represented by the attribute CLAUSALITY defined for phrases and set by the *clausality_principle*.

(32)



In the current version of the ALE grammar, the most frequent types of relative clauses, namely noun modifying ones, are implemented (*wh-rel* and *mark-rel* types). The *wh-rel* clauses can begin with nominal relative pronouns, e.g., *który* (*who/what) , jaki (which), kto* (*who), co* (*what),* (34a), or adverbial relative pronouns, e.g., *gdzie* (*where*), *skąd* (*where from*), (33b). *Mark-rel* clauses begin with the relative marker *co*, (31).

(34) a. Idzie chłopiec$_{nom,sg,m1}$, który$_{nom,sg,\ masc}$ __$_{nom}$ kupił Marii kwiaty.
goes boy who bought Mary flowers

[4] The problem of relative clauses is carefully discussed in Mykowiecka (2000,2001).

[5] The relation max-one says that lists 1 and 2 can have only one element in sum.

‘Here comes the boy who bought Mary flowers.’

b. tam, gdzie rosną poziomki
‘there, where strawberries grow’

The implementation does not cover relative clauses which modify entire sentences and free relatives, (35).

(35) a. Jan tańczył, co się nie często zdarza.
John danced, what self not often happen.
‘John danced, what happened not often.’

b. Weź co chcesz.
‘Take what you want.’

Coordination

Coordination is a very common phenomenon in natural languages. There are different types of coordination. It is possible to coordinate phrases of the same categories but also of different categories. Coordination may apply to constituents but coordination of non-constituents or partial constituents (phrases which share arguments) is also quite common in natural languages.

Our analysis of coordination in Polish is based on the HPSG account of constituent coordination presented in Paritong (1992). As a consequence, the conjunction is treated as a functional head with coordinated phrases as its complements. The formal analysis comprises[6] several types of coordination, e.g. constituent coordination, see (36a), where the subject *Jan* (*John*) and the complement *obraz* (*a picture*) is shared by the two verbs: *namalował* (*painted*) and *sprzedał* (*sold*). We can also deal with coordination of different categories but only in the case of modifiers, see (36b).

The implementation, due to low efficiency, had to be simplified and only basic ideas of the theoretical account have been incorporated.

(36) a. Jan namalował i sprzedał obraz.
‘John painted and sold a picture.’

b. Jan kupił dom tanio i na raty.
‘John bought a house cheap and by instalments.’

Conclusions

In this paper we have presented some aspects of the implementation of the HPSG Polish grammar fully described in Przepiórkowski at al (2002). The formal account is straightforward and consistent. The aim of the presented work was to check if it is implementable. Our goal was to develop an implementation as close to the linguistic theory as possible. Due to differences between ALE and HPSG, the implementation differs from the theoretical analysis in many respects but the main ideas of the theory have been preserved. It turned out, however, that the straightforwardness of the theory does not hold together with the effectiveness of the implementation. To keep the parse time reasonable, we had to simplify the implementation of several linguistic phenomena, so some of the subtheories are implemented only partially. Nevertheless, the main goal of out work was achieved and the implementation confirmed the correctness of the main theoretical ideas.

[6] The problem of coordination in Polish is also discussed in Kupść at al. (2000a, 2000b).

Bibliography

Bański, P. and Przepiórkowski, eds, (2000), *Proceedings of the First Generative Linguistics in Poland Conference*, Warsaw, Institute of Computer Science, Polish Academy of Sciences.

Bolc, L. (2002) Formal description of natural languages: An HPSG grammar of Polish. To appear in *Festschrift volume for Jorg Siekmann*, Lecture Notes in Artificial Intelligence, Springer-Verlag, Heidelberg.

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

Bouma, G., Malouf, R., and Sag, I. A. (2001). Satisfying constraints on extraction and adjunction. *Natural Language and Linguistic Theory*, 19(1), 1-65.

Carpenter, B. and Penn, G. (2001). *The Attribute Logic Engine (Version 3.2.1). User's Guide*. Carnegie Mellon University, Pittsburgh.

Czuba, K. (1995). *Zastosowanie dziedziczenia do analizy wybranych aspektów języka polskiego.* Master's thesis, Uniwersytet Warszawski, Warszawa.

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

Kupść, A., Marciniak, M., and Mykowiecka, A. (2000a). Constituent coordination in Polish: An attempt at an HPSG account. In Bański and Przepiórkowski (2000), pages 104-115.

Kupść, A., Marciniak, M., Mykowiecka, A., and Przepiórkowski, A. (2000b). Składniowe konstrukcje współrzędne w języku polskim: Próba opisu w HPSG. Technical Report 914, Institute of Computer Science, Polish Academy of Sciences.

Marciniak, M. (2001). *Algorytmy implementacyjne syntaktycznych reguł koreferencji zaimków dla języka polskiego w terminach HPSG*. Ph. D. dissertation, Polish Academy of Sciences, Warsaw.

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

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

Mykowiecka, A. (2002). Który vs. kto – Polish noun modifying relative clauses. In Zybatow et al., pages 127-336.

Mykowiecka, A. (2001). Polish Relatives with the marker *co*. In Przepiórkowski and Bański (2001), pages 124-134.

Paritong, M. (1992). Constituent coordination in HPSG. Technical Report CLAUS 24, Universität des Saarlandes, Saarbrücken.

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

Przepiórkowski, A., A. Kupść, M. Marciniak and A. Mykowiecka. (2002). Formalny opis języka polskiego. Teoria i implementacja. Akademicka Oficyna Wydawnicza RM, Warszawa.

Przepiórkowski, A. and P. Bański, editors, (2001), *Generative Linguistics in Poland. Proceedings of the GLiP-2 Conference*, Warsaw, Institute of Computer Science, Polish Academy of Sciences.

Przepiórkowski, A. (1999). *Case Assignment and the Complement-Adjunct Dichotomy: A Non-Configurational Constraint-Based Approach*. Ph. D. dissertation, Universität Tübingen, Germany. Available from: http://www.ipipan.waw.pl/mmgroup/ap-papers.html.

Przepiórkowski, A. (2000). ARG-ST on phrases headed by semantically vacuous words: Evidence from Polish. In D. Flickinger and A. Kathol, editors, *Proceedings of the 7th International Conference on Head-driven Phrase Structure Grammar*, pages 267-284. CSLI Publications, Stanford, CA.

Przepiórkowski, A. and Kupść, A. (1997). Verbal negation and complex predicate formation in Polish. In R. C. Blight and M. 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.

Przepiórkowski, A. and Kupść, A. (1999). Eventuality negation and negative concord in Polish and Italian. In Borsley and Przepiórkowski (1999), pages 211-246.

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

Sag, I. A. (1997). English relative clause constructions. Journal of Linguistics, 33(2), 431-483.

Szpakowicz, S. (1983) *Formalny opis składniowy zdań polskich.* Wydawnictwa Uniwersytetu Warszawskiego, Warszawa.

Świdziński, M. (1992) *Gramatyka formalna języka polskiego*, Wydawnictwa Uniwersytetu Warszawskiego, Warszawa.

Zybatow, G., U. Junghanns, G. Mehlhorn and L. Szucsich, editors, (2002) *Current Issues in Formal Slavic Linguistics*. Frankfurt/Main: Peter Lang.