Charles University

Faculty of Mathematics and Physics

AMBIGUITY OF PREPOSITIONAL GROUPS AND THE POSSIBILITY OF ITS AUTOMATIC PROCESSING

Summary of Doctoral Thesis

I-3 Mathematical Linguistics

Markéta Straňáková

Prague, 2001

INTRODUCTION

The objective of the present work is to analyze ambiguity of prepositional groups (Pg's) on the background of dependency based Functional Generative Description (FGD). Several basic types of Pg-ambiguity are presented and the possibilities of searching for their occurrences in the Prague Dependency Treebank (PDT) is illustrated. I aim at a formulation of linguistically-based rules for the detection of the appropriate sentence member (or members if two or more analyses are adequate) on which the given Pg can depend.

A general method of analysis suitable for automatic processing, an analysis by reduction, is introduced. I investigate four types of criteria for automatic detection of ambiguity of prepositional groups: (1) Criteria based on surface word order and (2) criteria exploiting valency frames of verbs, nouns and adjectives; optionally (3) rules based on semantic features can be used. The fourth type, criteria concerning the word order positions as combined with the relationship of a particular modifier to other modifiers, serves for excluding inappropriate structures. A stepwise application of these criteria is proposed following the basic principles of the analysis by reduction. I discuss the usefulness of these criteria and the reliability of analyses obtained.

In the last part I show the possibility of an application of criteria proposed within a concrete automatic parsing system.

I. MOTIVATION

The problem of ambiguity of prepositional group belongs to the difficult problems of any automatic procedure of syntactic analysis in any framework (be they phrase structure oriented or dependency based). From a strictly syntactic point of view a Pg can modify (depend on) any verb, adjective (typically deverbal) or noun. The wide range of linguistic phenomena concerned can be illustrated by example (1).[1]

Example:

(1) *V období, kdy prudce **poklesl zájem** na domácím trhu, dokázala továrna část výroby exportovat.* (Prague Dependency Treebank, bmd03zua.fs #4, shortened)
[At – time – when – sharply – fell – demand – on – domestic – market – managed – factory – part – (of) production – (to) export.]
(At the time when the demand on the domestic market sharply fell, the factory managed to export part of its production.)

The Pg *na (domácím) trhu* can be treated either (a) as a valency modifier (Patient) of the noun *zájem*, or (b) as a free modifier (Locative) belonging (ba) either to the verb *poklesnout*, or (bb) to the noun *zájem*. (The valency slot of Pat with *zájem* is occupied by a General Participant in cases (ba) and (bb).)

In general a sentence with a Pg that can be analyzed as a modifier belonging to two (or more) distinct sentence members can have two (or even more) different meanings (as in (a) and (ba)) – thus the need for criteria that allow to detect the occurrences of ambiguous Pg's.

The possible syntactic functions of prepositional groups in Czech were studied in the seventies, when a partial algorithmic procedure for the detection of several types of adverbal

---

[1] Notation: Here and in the sequel, the Pg's under discussion are underlined, and their possible governors are are printed in bold face.

and adnominal modifiers was formulated (only a small part of these algorithms was published, some results are presented in Panevová & Sgall (1979), (1980) and Bémová (1979)). At present, the existence of rich Czech corpora, the Czech National Corpus and the Prague Dependency Treebank, allows to study the problem of ambiguity of Pg's from broader perspectives.

Most of the reported approaches to disambiguation of Pg's in other languages, esp. in English, are statistical or corpus-based, and they usually try to solve a rather simplified problem: they consider only prepositions whose attachment is ambiguous between a preceding noun phrase and a verb phrase (Ratnaparkhi (1998)). My approach is 'linguistically' based, it aims at a formulation of linguistically-based rules for the detection of the respective sentence member (or members, in the case of ambiguity) on which Pg can depend. Such rules can have different priority, which supports the evaluation of particular syntactic structures.

II.   THEORETICAL BACKGROUND

The linguistic framework of my research is formed by the dependency based Functional Generative Description (FGD) of the Czech language, described e.g. in Sgall (1967), Sgall et al. (1986) and Sgall (1992).

The FGD can be classed with stratificational approaches – the description of language is divided into several levels. This is important for the notion of ambiguity – it is connected with asymmetry of the form of the sentence and its meaning.

Definition of ambiguity on the level n:

Ambiguity on the level n can be defined as a relation between two (or more) units on the level n+1 ('functions') and their common expression ('form') on the level n (Panevová (1981)).

I distinguish three types of ambiguity – **syntactic ambiguity** (concerning two levels of syntax, underlying and surface), **morphemic ambiguity** (structural ambiguity, between the level of (surface) syntax and morphemic level) and **phonological ambiguity**.

I use the term **prepositional group** as a parallel of the term prepositional phrase known from the phrase structure oriented grammars (it seems to be better to avoid using the latter term due to its connotations). In the simplest case, a Pg is constituted by a preposition and a noun in a relevant morphemic case ('core' noun of Pg). A more complex Pg may be simplified by means of an analysis by reduction (which will be introduced later) in order to obtain the former case (so called reduced Pg).

Presupposing full morphological analysis, the ambiguity relevant for Pg's can be sorted into two groups:

• From the purely syntactic point of view the **morphemic (or structural) ambiguity** is characteristic for prepositional groups. Pg may in principal modify:

◦ any noun preceding the tested Pg in surface word order; any verb form being understood as the left-hand boundary (the exceptions are discussed below);

◦ any verb (autosemantic verb, modal verb as well as copula);

◦ any adjective (again, any verb form serves as a boundary).

The morphemic ambiguity (with unique morphological) information is systematic for the sentences with Pg's, it is determined by general syntactic rules.

• As some prepositions govern two (exceptionally three) morphemic cases (as e.g. *na* [on] or *v* [in], both requiring either Accusative or Locative) **phonological ambiguity** can also appear in connection with Pg's. The occurrence of phonological ambiguity is only accidental, inherent in the forms of concrete lexical items.

Whereas the potential morphemic ambiguity is resolved during syntactic analysis, the possibility of phonological ambiguity is detected earlier, during morphological (pre)processing – the lexical item in consideration (the 'core' noun in Pg, in our case) bears ambiguous morphemic information. In such cases, syntactic analysis splits into several branches in which all possible combinations of the input information are processed, i.e. each combination of possible morphological information is treated separately. If two (or more) branches of syntactic analysis (with distinct Pg's) are successful, then phonological ambiguity is stated as relevant for the given sentence, i.e. as a source for its structural ambiguity.

This is the reason why we focus on the morphemic ambiguity and on the proposal of criteria for its detection, the phonological ambiguity is treated implicitly.

III. AUTOMATIC SEARCHING FOR 'SUSPICIOUS' SYNTACTIC STRUCTURES

I use the syntactic annotations from the Prague Dependency Tree Bank (PDT), annotation on so called analytical level (for description see e.g. Hajič (1998)), as the basic source of data for Czech. Three 'suspicious structures' are defined with respect to Pg-ambiguity, i.e. syntactic structures which can signalize the occurrence of Pg-ambiguity.

Definition: (basic 'suspicious structures' with respect to Pg-ambiguity)

(a) a syntactic structure where a verb or a noun (V/N) is modified by another noun (N_g) which is modified by the tested Pg (i.e. (V/N (N_g (Pg)))) in the linearized form);

(b) a syntactic structure where a noun (N) is modified by an adjective (Adj) which itself is modified by the tested Pg (i.e. (N (Adj (Pg))) in the linearized form), the whole subtree modifies a verb (V);

(c) a syntactic subtree where a verb or a noun (V_g/N_g) is modified by the tested Pg; another noun (N) appears as a brother of this Pg (i.e. (V_g/N_g (N Pg)) in the linearized form).
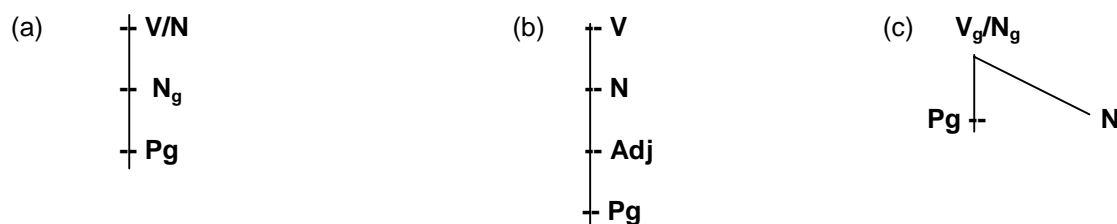
(a)
```
V/N
|
N_g
|
Pg
```

(b)
```
V
|
N
|
Adj
|
Pg
```

(c)
```
        V_g/N_g
       /      \
    Pg          N
```

Figure: Basic 'suspicious structures' with respect to Pg-ambiguity

Macros for searching for 'suspicious sentences':

I have designed two macros for Graph editor (i.e. editor used for the purposes of PDT) which make it possible to search for sentences whose syntactic structure contains one of the pre-defined subtrees – macro A covers the a-type and the b-type structures and macro B deals with the c-type structure.

Note:

These are only the basic types of 'suspicious structures'. The macros are able to detect also more complicated situations:

• they handle coordination in any position of the defined schemata;

• all nouns in the pre-defined structures can appear without or with a preposition (which has impact on the shape of analytic tree);

• macro A searches for all occurrences of any verb or noun laying on the branch (of syntactic tree) from the root of the tree to the node representing the tested Pg;

• macro B examines all brothers of the node representing the tested Pg.

At the first stage of my inquiry I focussed on Pg's with prepositions *na* [on] and *v* [in] (both with Accusative and Locative). A sample of 1250 sentences from PDT was tested and about 150 sentences with 'suspicious structures' were received for each preposition. (Note that different types of 'suspicious structures' often combine in one sentence.)

A question arises – why do we use a syntactically annotated corpus rather then only morphologically annotated texts? Prepositional groups prototypically satisfy word order constraints (which are described in part IV.1.), clear boundaries in searching for potential Pg governors can be stated. This fact supports the usefulness of morphologically tagged corpus. However, some linguistic constructions exist where these constraints fail. Then the syntactic structure may allow to detect possible Pg governors. As I want to describe those linguistic phenomena where it is useful to relax the word order constraints for Pg modifiers (and to permit non-projective structures), the syntactically annotated corpus seems to be a good resource.

The same reasons had led to the decision not to reflect surface word order in the macros. This decision makes it possible to capture also the occurrences of the Pg-modifiers in non-typical word order positions. (With respect to the basic aim of the macros – to collect source data for a manual analysis – the fact that also inappropriate 'syntagms' are found is not crucial, the amount of the relevant data obtained is more important.)

IV.   PROPOSAL OF THE CRITERIA

Before introducing criteria proposed for the detection of Pg-ambiguity (and for Pg-disambiguation, as the case may be) the prerequisites for their application must be presented.

Assumptions:

• The system of syntactic and semantic rules proposed supposes the existence of a lexicon with a detailed information on the valency slots of verbs, nouns and adjectives. Moreover, the theoretically based notion of valency is extended in favour of the aim of automatic syntactic analysis – morphemic forms for modifications commonly used with a particular lexical item are listed in its frames too, being called quasi-valency.

• Verbonominal collocations, i.e. frozen collocations of (primarily) verbs and nouns, are included in the lexicon, together with their valency (optionally).

• Semantic features of nouns are stored in the lexicon as well as the possible semantic functions of prepositions (optionally).

• The rules proposed here are supposed to operate on the results of the morphological analysis; the analysis of compound (analytical) verb forms, elementary noun groups and frozen collocations is presupposed, too.

Basic Strategy

The basic strategy of the approach to searching for relevant criteria for analysis of prepositional groups, consists in the **method of a stepwise refinement of rules** proposed for the detection of ambiguity of prepositional groups, or for their disambiguation.

1. In the first step only relatively simple rules are applied, especially rules based on typical word order and basic rules exploiting valency information.

2. Refinement of this 'core' analysis follows, the word order criteria are relaxed in defined cases which may lead to non-projective structures. Richer valency information is exploited including e.g. information of word collocations (with their valency) or semantic features of particular words (if they are at disposal).

3. The phase of 'pruning' follows which is characterized by complex rules leading to an exclusion of inappropriate non-projective structures established in the preceding steps.

The structures obtained are evaluated according to the type of rule used – this in some way corresponds to their 'reliability'.

Method of Analysis – Analysis by Reduction (RA)

The general framework of the method adopted consists of the analysis by reduction which can be seen as a model of an intuitive approach to analysis used by human.

General principles of the analysis by reduction:

• The analysis by reduction consists of **stepwise simplification** of a full sentence so that its syntactic (but only syntactic) correctness is preserved.

• An operation of **reduction** is a basic operation of RA: in each step the simplification is realized by deleting one word of the sentence and possibly rewriting other words.

• This process is **non-deterministic**, in each stage any of the mutually independent words can be deleted.

• The deletion is 'justified' by **linguistically-based rules**.

Application of RA to an automatic processing of Pg's:

• The analysis of natural language adopting the principles of analysis by reduction is realized by a sequence of steps in which the **surface realization** of the sentence is simplified.

• In each step **one sentence member is deleted**.

• The reduction of a Pg tested is 'justified' by **linguistically-based rules**.

• All **dependent** words must be deleted before their **governor** is deleted; otherwise linguistically based criteria 'justifying' reduction cannot be applied (The criterion for determining governor and dependent sentence members is introduced e.g. in Sgall (1998)).

• The possibility of **rewriting** is limited to changes in morphological categories of number, gender and case in order to preserve correctness.

• In each step of analysis based on reduction any of the mutually independent words can be deleted. The condition on (syntactic) correctness preserving is the only criterion for the choice of sentence members which can be deleted in a particular step. Thus the analysis of Pg usually splits into several branches.

• RA can be seen as a model of human analysis. Then the condition on **correctness preserving** is OK. For an automatic procedure the criterion of correctness must be weakened – the fact that the analysis is successful (that the sentence is reduced) 'justifies' particular branches of analysis.

• The linguistically-based rules have different reliability – based on this fact the analyses (dependency structures) obtained can be evaluated and ordered.

• Another principle, **principle of movement**, is used, which is important especially for investigation of word order criteria. The original sentences as well as their word order modifications are tested.

The following example illustrates the possibility of analysis based on reduction:

Example:

(2)   *Tyto rodiny mají **nárok** <u>na státní vyrovnávací příspěvek</u>.*
       (These families have a claim to a state compensatory allowance.)

Let me assume (for the sake of simplicity) that the congruent attributes *tyto* [these], *státní* [state] and *vyrovnávací* [compensatory] have been deleted in the first stages of the analysis (they are mutually independent). Thus, we obtain the following sentence (where only the phenomena important for our explanation are kept):
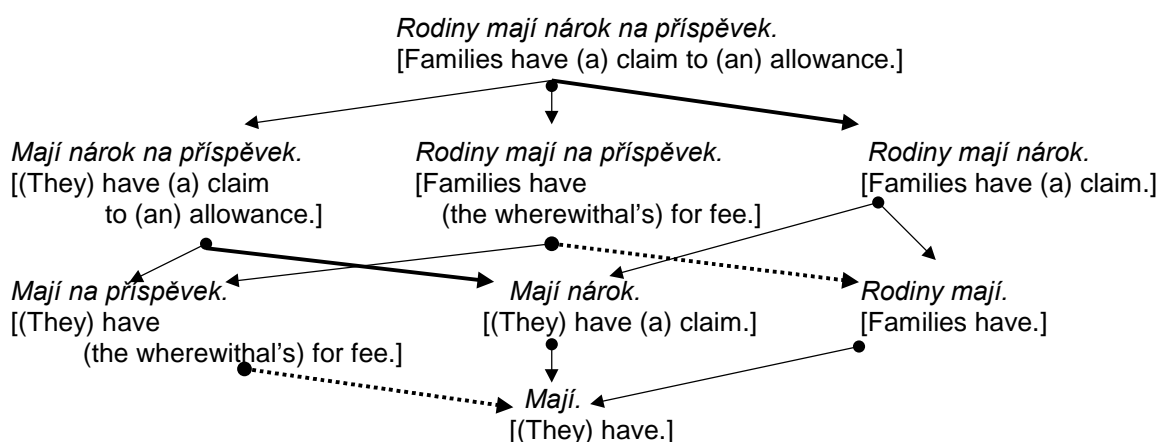

Figure: The branches of analysis of the sentence (2)

The analysis splits into several branches. The reduction of the Pg *na příspěvek* starts in several stages, with respect to the types of information that can be taken into account.
(a) If the verbonominal collocation *mít nárok* [to have (a) claim] is taken into account the reduction of the Pg *na příspěvek* [to allowance] starts in two stages (marked by bold arrows). The dependency pair *nárok na příspěvek* [claim to allowance] is determined (in agreement with valency requirement of the noun *nárok*, with high preference assigned).
(b) If the verbonominal collocations are not considered the analysis *mít na příspěvek* is also allowed. The reduction of Pg then starts in further two stages (marked by dotted arrows), the Pg is treated as a free verbal modifier, with low preference.
All these analyses satisfy the condition on preserving syntactic correctness, both are successful. But the second structure *mít na příspěvek* accidentally has another meaning than the original sentence has. Its meaning is connected with a frozen collocation *mít na něco*

(prostředky/peníze/...) [to have the wherewithal's for fee]. (The source of this coincidence is connected with the lexical analysis, it cannot be covered by general rules.)

It can be seen that the fact that dependency structures can be evaluated has a great impact. It allows to stress particular analyses and to sideline other analyses.

Naturally, the more information can be used, the better analysis is obtained. However, despite of the incompleteness of the input information the automatic procedure can grant a satisfactory answer to our problem. This is very important with respect to the fact that the input information is always incomplete, the knowledge involved in the natural language understanding (knowledge available for human) cannot be fully incorporated in any automatic system.

The goal of the work is to propose an analysis of Pg following the principle of reduction. It is realized as a sequence of criteria and the dependency pairs obtained are evaluated – their 'reliability' is given by the type of criteria used.

VI.1. Word Order Criteria

(Word Order Patterns, WOP's, and Word Order Constraints)

The surface word order serves as the basic purely syntactic clue for the Pg-ambiguity detection. Basic unambiguous WOP's and suspicious WOP's are defined with respect to Pg-disambiguation (concerning the surface representation of the sentence), WOP with preferable readings can be also stated.

Basic unambiguous WOP's:

• V (or N at the beginning of a clause) immediately followed by a Pg, **VPg type** (or **NPg type**, respective) (such V or N is the Pg governor);

• Pg at the very beginning of a clause, **PgV type** (V is the prototypical Pg governor in this case);

• Adj and Pg are separated by any verb form, **AVPg** and **PgAV types** (V is the Pg governor in such case).

If one of these WOP's is met then a single <u>dependency pair is created</u>, with low preference, preference of free modifiers.

Basic suspicious WOP's:

• A sequence of nouns at the beginning of a clause, **NNPg type**;

• Pg between a noun and a verb (from the left to the right), **VNPg type**;

• a sequence of a verb, a noun and Pg (in this order), **NPgV type**;

• Pg between a verb and an adjective, **VPgA type**, or Pg being followed by an adjective and a verb (in this order), **PgAV type**.

There can be strings of nouns instead of a single N in all of these word order patterns, all of them either without a preposition or in a prepositional case.

If one of these WOP's is met then <u>all possible dependency pairs are created</u>, with the same low preference of free modifiers.

WOP with a preferable reading:

• A noun being followed by a congruent adjective and Pg, in this order, $N_i A_i Pg$ **type** (then Pg is preferably treated as belonging to such Adj);

• Pg between a noun and a congruent adjective (from the left to the right), N_iPgA_i and A_iPgN_i **types**.

There are some regular exceptions to this rule as the specific contrastive position of Adj or collocations characteristic of scientific terminology.

If one of these WOP's is met then the <u>dependency of Pg on Adj is established</u>, with the highest preference.

Example:

(3) *Dívka* **rovná** <u>*na ramínku*</u> **vystavený** *kabát.* ([Panevová])
[girl – arranges – on – hanger – exposed – coat]
(The girl arranges an exposed coat on a hanger. / The girl arranges a coat exposed on a hanger.)

(3a) *Vystavený kabát* **rovná** <u>*na ramínku*</u>
[exposed – coat – arranges – on – hanger]
((She) arranges an exposed coat on a hanger.)

(3b) **Rovná** *vystavený* **kabát** <u>*na ramínku.*</u>
[arranges – exposed – coat – on – hanger]
((She) arranges an exposed coat on a hanger.)

In (3) the surface word order agrees with **VPgA** type, the Pg *na ramínku* [on a hanger] appears between the verb *rovnat* [to arrange] and the adjective *vystavený* [exposed]. Two dependency pairs are established, *rovnat na ramínku* [to arrange on a hanger] and *vystavený na ramínku* [exposed on a hanger], which is in agreement with the meaning of the sentence

(3a) is unambiguous, a single structure is created, *rovnat na ramínku* [to arrange on a hanger] (based on WOP **AVPg**).

The (3b) variant is interesting – if only projective structures are allowed then the 'suspicious' WOP, namely the **VNPg** type is relevant, the dependencies *rovnat na ramínku* [to arrange on a hanger] and *kabát na ramínku* [coat on a hanger] are created. If also non-typical positions of Pg are tested (described below), also the dependency pair *vystavený na ramínku* [exposed on a hanger] must be admitted.

Non-typical Positions of Pg's

Prototypically, Czech constructions with Pg's meet the constraints defined above. However, there are some constructions in Czech where these conditions must be relaxed (for the sake of the adequacy of the resulting analysis).

I distinguish three such constructions:

• **verbonominal collocations**, i.e. collocations of a noun and a verb (often with very general meaning) where the noun part has (usually) a valency position filled in by a Pg;

• **nominal modifiers** which satisfy valency requirements typical for verbs (with the only additional condition that in the given sentence the tested Pg cannot be treated as a verbal valency modifier);

• **non-projective Pg-modifiers of adjectives,** namely Pg-modifier which is separated from its adjective governor by a noun modified by this adjective, A_iN_iPg **type**.

Verbonominal collocations represent the constructions where it seems to be adequate to relax the word order constraints. The fact that such a Pg is treated as a nominal modifier (in

agreement with the valency requirements) may lead to a non-projective structure (which seems to have the same degree of reliability as a prototypical structure with word order constraints satisfied, as in the following example.)

Example:

(4) *Na uveřejnění odpovědi* **má** *podle něho každá osoba* **právo** *...* (PDT, bmd23zua.fs #46, modified)

[on – publication – (of) answer – has – according to – him – every – person – right]
(According to him every person has right to the publication of his/her answer.)

The noun *právo* [right] – constituting a verbonominal collocation together with the verb *mít* – has a valency slot requiring *na*+Acc modifier; the Pg *na uveřejnění (odpovědi)* [to publication (of his/her answer)] meets this requirement creating a non-projective construction.
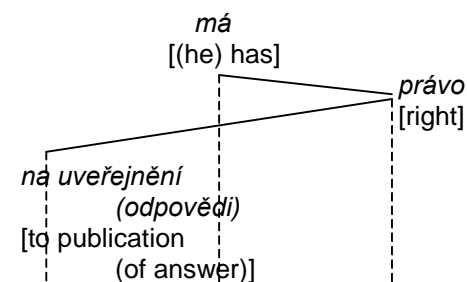

Figure: Non-projective dependency tree of the collocation *mít právo (na něco)*

We have adapted the strategy of 'careful' relaxation of the word order constraints, i.e. the relaxation is allowed only with well described phenomena in order to reach a linguistically appropriate solution. Each inadequate relaxation leads to an explosion of analyses obtained by an automatic procedure, most of them inappropriate.

VI.2. Valency Frames

The valency frames of verbs, nouns and adjectives play a crucial role for the Pg-disambiguation. The extension of this theoretically based notion in two directions seems to be useful:

• The notion of valency primarily pertains to the level of underlying representation of a sentence (e.g. Panevová (1974, 1975), (1994), (1998) and (2000)). However, the valency frames can be fruitfully interpreted also with regard to the means of their expression in the surface structure. One feature of such interpretation is important from the point of view of automatic processing. In the **surface (morphemic) form** of the sentence (almost) any member of valency frame is deletable (at least in the specific contexts as e.g. a question-answer pair). Thus, the knowledge of the valency frame of a particular item can be used in 'one direction' only. If a member (e.g. a Pg) satisfying some valency requirement is present in the sentence, then it can be treated as such a valency modifier. Its absence, on the contrary, does not mean that the sentence is incorrect.

• Two types of valency information (concerning the level of underlying representation, both stored in the lexicon) are to be distinguished: the valency as it is described in theoretical studies and the quasi-valency – which may be paraphrased as 'commonly used modification' of a particular item – proposed for the purposes of parsing.

I discuss and exemplify the usefulness of valency information of particular lexems (verbs, nouns and adjectives) as well as the valency characteristics of word collocations

namely verbonominal collocations. The discussion results in the following classification of the reliability of valency frames.

'Reliability' of Valency Information

Generally the chance that the sentences with a Pg depending on different governors – in agreement with different valency requirements – share the same meaning is based on the cognitive content (on inferencing). This results in the necessity of preserving all structures satisfying valency requirements of any unit within a parsing procedure, with the only constraint for modifiers of adjectives (any verb form serves as a block for the attachment of Pg to an adjective).

It seems to be useful to classify different types of valency requirements (taking into account the grammatical – not just textual – omissibility of the dependents):

Evaluation of valency modifiers:

 • The 'classical' valency frame of verb has the highest priority.

 • The valency requirements of adjective follow (if the word order restrictions are fulfilled and the 'weight' is not strengthened by the WOP with a preferable reading).

 • The valency frame of noun (with prescribed form) succeeds.

 • The quasi-valency frame of a particular POS is of a lower reliability:

   ◦ the quasi-valency frames of nouns;

   ◦ the quasi-valency frames of verbs and

   ◦ the quasi-valency frames of adjectives.

This classification of valency requirements serves for the evaluation of the final syntactic structures of a sentence with their overlappings, and thus for the classification of the structures according to their reliability.

The following examples illustrate the situations where there are two structures fulfilling the valency requirements of different priorities. All analyses satisfying valency or quasi-valency requirements of particular words in the analyzed sentence must be preserved in such cases.

Example:

(5) ***Zákon*** *na ochranu nájemníků* **dbá** *snad až příliš.*   (based on PDT, bl101js.fs #18)
[law – on – protection – (of) tenants – looks after – perhaps – even – too.]
(Perhaps the law looks only too much after the protection of the tenants. / Perhaps the law for the protection of the tenants looks only too much after it.)

Both the verb *dbát* [to look after] and the noun *zákon* have a valency, resp. quasi-valency position which can be satisfied by tested Pg *na ochranu (nájemníků)* [for the protection (of the tenants)]:

   $dbát_1$ [to look after]…   Act | Pat (*na*+Acc / *o*+Acc)[2]

   zákon [law]   …   Pat (*o*+Loc)

   zákon [law]   ...   ^Aim (*na*+Acc)[3]

---

[2] There is also another frame of the verb *dbát*, with Patient in genitive case :
     $dbát_2$ [to look after] … Act | Pat (Gen)
[3] The symbol ^ denotes a quasi-valency modifier in the sequel.

The Pg *na ochranu (nájemníků)* is analyzed in two ways. Either as the valency modifier of the verb *dbát* (filling in the Patient slot, 'high priority' according to the classification proposed), or as the quasi-valency modifier of the noun *zákon* (as *zákon* requires Aim modifier with the form *na*+Acc, 'lower priority').

It is obvious that the valency requirements of verb must be preferred. However, there are contexts in which adnominal analysis is relevant (e.g. as the answer to the question *Dbají na tato omezení zákony?* [Do the laws look after these limits?].

Example**:**

(6)   *Japonský **lék** na kanadské problémy **nepomohl**.*
      [Japanese – medicine – on – Canadian – problems – did not help]
      (The Japanese medicine failed with Canadian problems. / The Japanese medicine for Canadian problems met with failure.)

Both the noun *lék* [medicine] (a primary noun, its valency frame consists of an optional Patient modifier) and the verb *pomoci* [to help] (which has quasi-valency modifier in the relevant meaning) have positions in their frames which can be satisfied by the tested Pg *na (kanadské) problémy* [on Canadian problems]:

      lék [medicine]   … Pat (*na*+Acc/*proti*+Dat)

      pomoci$_2$ [to help]   …   Act | ^Pat (*na*+Acc)[4]

The sentence actually exhibits a structural (morphematic) ambiguity – the tested Pg can be treated either as a non-congruent attribute of the noun *lék* [medicine] or as a quasi-valency modifier of the verb *pomoci* [to help].

Both structures will be established, the former with high preference (as satisfying valency requirements), the latter with lower preference (as a quasi-valency modifier). This fact is important because they have different truth conditions (and hence different meaning), as the possible continuations show:

(6a)   *Japonský lék na kanadské problémy **nepomohl** (Kanadě, pomohl však na problémy Británie.)*
      (The Japanese medicine did not help Canada with Canadian problems, but it helped with the problems of Britain.)

(6b)   *Japonský **lék** na kanadské problémy nepomohl (Británii, přestože Kanadě pomohl.)*
      (The Japanese medicine for Canadian problems did not help Britain, though it helped Canada.)

In (6a) the analysis *(ne)pomoci na (kanadské) problémy* [(not) to help with Canadian problems] is adequate. On the other hand, in (6b) the different analysis *lék na (kanadské) problémy* [medicine for Canadian problems] seems to be proper.

The difference between the two analyses obtained seems to be clearer in the sentences (6a') and (6b') which have the same syntactic structures as (6a) and (6b) have:

(6a')   *Lék na kašel **nepomohl**, ale teplotu srazil.*
      (The medicine did not help a cough, but it lowered the temperature.)

(6b')   *Lék na kašel nepomohl, spíš mu pomohlo, že zůstal v posteli.*
      (The medicine for a cough did not help, rather it helped that he stayed in bed.)

---

[4] The valency frame of the verb *pomoci*:
      pomoci$_1$ [to help] … Act (Pat (*s*+Ins/inf/klauze) Addr (Dat)

The application of valency information is a relatively very safe criterion for the detection of Pg-governors. Nevertheless, it presupposes very detailed and extended linguistic research.

VI.3. Semantics

It is clear that the capability of the human to analyze a sentence is based to a great extent on the understanding of its meaning. But the semantic and even pragmatic aspects involved in the natural language understanding cannot be incorporated in their complexity in any automatic system. Thus before building any syntactic parser it is necessary to make some decisions concerning the types of input information the parser will take into account. There are parsers using certain types of semantic features (e.g. Oliva (1996)), on the other hand, some approaches exist the goal of which is to capture a pure surface syntactic structure of a sentence. The Robust Parser for Czech (Kuboň (1999)) can be treated as an example of such syntactic approach using only limited information about particular lexical items with restricted valency frames.

Since a lexicon for Czech containing information about the semantic features within the noun items is being developed for the purposes of other projects (for the proposal of particular lexical items see Skoumalová (1994)) it seems to be useful and efficient to use these features as supplementary criteria for the solution of the Pg-ambiguity.

Criteria for semantic features are given e.g. in Buráňová (1980). The author proposes a method for semantic classification of nouns based on their possible functions on the level of underlying representation. These semantic features are used especially for identifying free modifiers, obligatory ones as well as optional ones. I discuss a possibility of using slightly modified semantic features of nouns together with semantic features of verbs and prepositions for the purposes of Pg-disambiguation.

If one of the mentioned valency frames contains some free modifiers, especially temporal and local, the use of semantic criteria even for the identification of these valency modifiers is obvious.

If the Pg in a particular sentence is identified neither as a valency nor as a quasi-valency modifier of some verb, noun or adjective, then it is time to:

• use semantic features of all potential Pg-governors in the sentence, the semantic characteristics of the preposition in the Pg and the semantic features of the 'core' noun in the Pg.

The following example illustrates the application of semantically based rules for chains of nouns and prepositional groups with prepositions *na* [on] and *v* [in] (primarily prepositions with local meaning, together with preposition *do* [to]). I formulate rules for nouns indicating real objects (bearing the feature 'concrete'), and nouns with features 'time' or 'local' – the relevant rules are the most evident. Their contribution is apparent especially in sentences with several local or temporal free modifiers.

Example:

(7) *Osmnáctiletý mladý muž s nožem v ruce* **přinutil** *v pátek v poledne na zastávce na Jírově ulici v Brně řidiče MHD, aby ho odvezl do Opavy a zpět.* (PDT, bl209pp.fs #32, modified)

[eighteen years old – young – man – with – knife – in – hand – forced – in – Friday – in – noon – on – stop – on – Jírova – street – in – Brno – driver – (of) MHD – that – him – take – to – Opava – and – back]

(On Friday noon on (bus)stop on Jírova street in Brno an eighteen years old youngster with knife in his hand forced the driver to take him to Opava and back.)

The sentence is highly ambiguous – there are five free modifiers the governors of which cannot be determined by means of word order and valency criteria. The semantically based rules allow to detect the most probable reading (from tens of possible analyses).

The Pg's *v pátek* [on Friday] and *v poledne* [on (the) noon], both with temporal meaning, are usually syntactically independent, thus they modify the verb *přinutit* [to force].

If temporal modifiers precede local modifiers then they are typically independent (this can be understood as an impact of systemic ordering introduced e.g. in Sgall et al. (1986)). Thus the Pg *na zastávce* [on (bus)stop] is also analyzed as a verbal modifier (dependency pair *přinutit na zastávce* [to force on (bus)stop] is created).

There is a noun *zastávka* [(bus)stop] with the semantic feature **'local'** which is followed by the Pg *na (Jírově) ulici* [on (Jírova) street] bearing semantic features **'local'** plus **'proper name'**, thus the Pg is preferably treated as a non-congruent attributive modifier of preceding noun, e.g. the analysis *zastávka na (Jírově) ulici* [(bus)stop on (Jírova) street] is preferred.

If the information is available that *Jírova ulice* takes place in *Brno* (model of the world), then the Pg *in Brno* ('wider' determination) is preferably a non-congruent attribute of the 'closer' local identification *na (Jírově) ulici*.
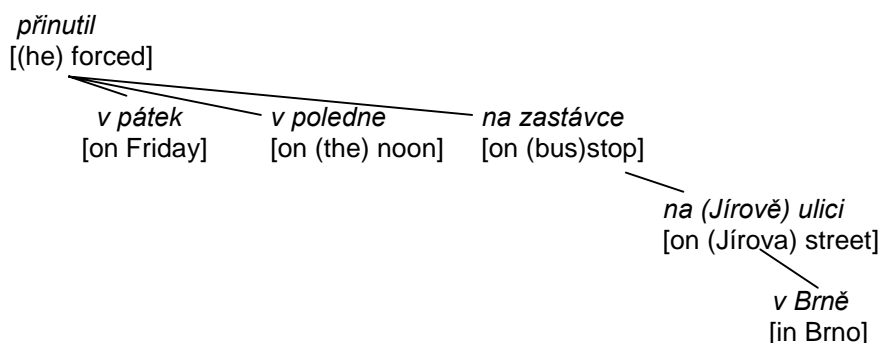

Figure: The only preferable structure of the sentence (7) based on semantic features

The semantically based observations allow to prefer one of tens analyses of sentence with a number of temporal and local modifiers.

Discussing the examples of rules based on semantic features of particular lexical items I must conclude that – despite the encouraging results for particular sentences – the possibility of exploitation of such type of semantic information for Pg-disambiguation is disputable and one must be very careful if s/he wants to use it. The semantic features are suitable for identifying free modifiers, especially temporal and local ones, and the meaning of individual prepositions may serve as a good tool for further specification of particular free modifiers. In case of morphemic ambiguity of prepositional groups the rules based on semantic features can be used only as supplementary criteria for evaluation of existing syntactic structures. They can help for ordering the structures and for setting up preferable readings of the examined sentence.

VI.4. Word Order and Formal Criteria

Firstly, prepositional groups in prototypical positions were investigated. Secondly, it was shown that language constructions exist, typically connected with valency requirements of nouns, where the word order constraints stated must be relaxed. Such relaxation may lead to non-projective structures. In the third phase I concentrate on relatively complex criteria which allow to eliminate inappropriate non-projective dependency trees.

The 'separation principle' serves as an example of such complex criteria.

'Separation principle':

If there is a chain consisting of a verb and three nouns or prepositional groups ($V – N_i/Pg_i – N_j/Pg_j – N_k/Pg_k$) in the surface representation of a particular sentence where the second N_j/Pg_j has been analyzed as a verbal modifier (participant or free one) then the third one N_k/Pg_k cannot be a daughter of the first one N_i/Pg_i (i.e. a noun or prepositional group as a verbal modifier cannot intervene between the preceding and the following noun or prepositional group).

The analysis of the following sentence shows the contribution of this principle.

Example:

(8)   *Vysoký činitel ... se snaží* **přiměl** *Palestince* **k ústupkům na poslední chvíli.**   (PDT, lc02zu.fs, shortened, modified)
    [Top – official   – … – tries – to force – Palestinians – (to) concessions – at – last – moment]
    (The senior official … tries to force the Palestinians to retreat at the last moment.)

Based on the valency requirement of the verb *přimět* [to force] which can be satisfied with the Pg *k ústupkům* [to concessions] the dependency pair *přimět k ústupkům* is established:

    přimět [to force]   …   Act | Pat (*k*+Dat/inf/clause-aby) Addr (Acc)

Now, the Pg *k ústupkům* [to concessions] serves as a member separating the first noun in the respective chain, *Palestinci* [Palestinians] (verbal participant, Addressee), and the third member, Pg *na (poslední) chvíli* [at (the) last moment]. According to the 'separation principle' the Pg *na (poslední) chvíli* [at (the) last moment] cannot be treated as an attribute of the first noun *Palestinci* [Palestinians] (despite of the semantic adequacy of such dependency pair). It can modify either the second Pg *k ústupkům* [to concessions] or the verb *přimět* [to force].

Analogously in the following modifications:

(8a) *Vysoký činitel ... se snaží* **přimět** *Palestince* **k odchodu** *z Jeruzaléma.*
    [Top – official   – … – tries – to force – Palestinians – to – departure – from – Jerusalem]
    (The senior official … tries to force the Palestinians to leave Jerusalem.)

 (8b) *Vysoký činitel ... se snaží* **přimět** *k odchodu* **Palestince** *z Jeruzaléma.*
    [Top – official   – … – tries – to force – to – departure – Palestinians – from – Jerusalem]
    (The senior official … tries to force the Palestinians from Jerusalem to leave (Jerusalem/another place).)

In (8a) the dependency *Palestinci z Jeruzaléma* is forbidden (the Pg *k odchodu* [to departure] serves as a separating member) – the Pg *z Jeruzaléma* [from Jerusalem] can be treated either as a verbal modifier, or as a modifier of the Pg *k odchodu* [to departure]. (The latter analysis

is preferred as the Pg *z Jeruzaléma* [from Jerusalem] can fulfill the valency requirement of the noun *odchod* [departure] for Direction modifier.)

On the other hand, in (8b) the analysis *odchod z Jeruzaléma* [departure from Jerusalem] is excluded (though the Pg *z Jeruzaléma* [from Jerusalem] fulfills valency requirements of the noun *odchod* [departure] as in (8a)), in accordance with the meaning of the sentence. The Pg *z Jeruzaléma* [from Jerusalem] can modify either the noun *Palestinci* [Palestinians] or the verb *přimět* [to force]. (The preference of the former analysis can be stated by means of semantically based rules.)

The 'separation principle' represents complex criteria on the basis of which one type of inappropriate non-projective structure (created in the previous phases of analysis) may be excluded. This field of research leading to formulation of other complex restrictions on the possible structure of sentence promises further interesting results.

V. ARRANGEMENT AND EVALUATION OF THE CRITERIA PROPOSED

I have already stated that my approach to automatic analysis is based on the notion of analysis by reduction. The analysis by reduction (RA) is realized as a sequence of steps in which particular criteria proposed are applied. They are applied in each branch of analysis, in each step in which a Pg can be reduced. These criteria are related either to the original sentence or to the (partially) reduced sentence.

The morphologically unambiguous input is presupposed (if phonological ambiguity appears in the particular sentence RA splits into several branches, the possible combinations of the input information are processed separately).

Arrangement of the criteria:

• All valency and quasi-valency requirements of verbs, nouns and adjectives are examined. The tested Pg potentially satisfies all such requirements (section IV.2.).

• Word order pattern (WOP) of simplified sentence is specified.

• Information following from word order pattern of simplified sentence is applied. The possible new dependencies are marked (section IV.1.), the dependencies already established by valency requirements are not doubled.

• The formal criteria are checked with regard to the original sentence. If any new structure violates these criteria then it is excluded from further analysis (section IV.4.).

• The possibility of the relaxation of word order constraints is checked (with respect to the rules discussed in IV.1.).

If verbonominal collocations are taken into account, the adverbal analysis is excluded (the adnominal one is preserved).

• If at least one new structure is preserved in the particular reduction step then Pg is deleted, analysis continues in the following step of this branch. Otherwise this branch of analysis fails, Pg is preserved, analysis continues in another branch.

If at least one branch of the analysis based on reduction is successful, then the whole analysis is successful.

Evaluation:

The dependency structures obtained are evaluated according to the type of criteria used for their establishment.

• If so called WOP with preferable reading for adjective is met in a sentence, then the highest preference is assigned to the structure with a Pg analyzed as an adjectival modifier, be it its valency or free modifier.

• The structures satisfying valency requirements follow, in the order introduced (Section IV.2.).

• The dependency pairs based on 'unambiguous' and 'suspicious' WOP's follow:

◦ rules using semantic features (if they are considered) may contribute to the evaluation of particular free modifiers;

◦ otherwise all structures based on 'unambiguous' and 'suspicious' WOP's have the same degree of preference.

• The lowest preference is assigned to a non-projective structure with a Pg treated as a modifier of an adjective separated from it by the nominal governor of the whole nominal group.

This evaluation is proposed for all structures with Pg under consideration, independently of their morphological characteristics. All dependency pairs obtained in all possible branches of RA (i.e. structures belonging to both types of ambiguity relevant for Pg's, structural (morphemic) as well as phonological) are compared.

Example:

(9) *Neschopnost způsobila skluz v realizaci autonomie **omezené** na území Golanských výšin*. (PDT, blc02zu.fs, shortened)
[Incapability – caused – delay – in – realization – (of) autonomy – limited – on – territory – of Golan – Highs.]
(The incapability caused a delay in realization of an autonomy limited to the territory of the Golan Highs.)

Let me focus on the Pg *na území (Golanských výšin)* [to/on (the) territory (of Golan Highs)] which bears ambiguous morphological characteristic, i.e. phonological ambiguity is stated. Analysis splits into two branches, the Pg is unambiguous in each of them (for the sake of simplicity, the reduced Pg *na území* [to/on (the) territory] is taken into account.

A. The Pg is analyzed as *na*+Acc, it satisfies valency requirement of the adjective *omezený* [limited]:

omezený [limited]   …   Act (Ins) Eff (*na*+Acc)

Again, the analysis of the Pg splits into several steps[5].

(a)   *… způsobila skluz v realizaci autonomie omezené na území.*
(...(it) caused a delay in realization of an autonomy limited to the territory.)

There is one valency requirement detected – *omezený na území* [limited to (the) territory]. The WOP of the type N_iA_iPg is specified, i.e. WOP with a preferable reading for adjectives. No new dependency pair is established based on WOP. No formal criterion is violated with respect to the original sentence. There are no reasons for the relaxation of word order

---

[5] Only some successful branches of analysis are presented – RA is non-deterministic, different branches may lead to the same structures.

constraints. One valid structure is obtained, Pg *na území* is deleted, the analysis continues with the next step, see (b). (This branch leads to successful analysis.)

(b)   *... způsobila skluz v realizaci autonomie na území.*
          (...(it) caused a delay in realization of an autonomy to the territory.)

No valency requirement is detected. The WOP of the type **VNPg** is specified. Four new dependency pairs are established based on WOP - *způsobit na území* [caused to (the) territory], *skluz na území* [delay to (the) territory], *realizace na území* [realization to (the) territory] and *autonomie na území* [autonomy to (the) territory]. No formal criterion is violated with respect to the original sentence. There are no reasons for the relaxation of word order constraints. Four valid structures are obtained, Pg *na území* is deleted, and the analysis splits into four branches and continues. (These branches lead to a successful analysis. There are no means available to control the adequacy of free modifiers.)

Five different analyses of the Pg *na území* [to (the) territory], *na*+Acc, were created during RA. One of them, *omezený na území* [limited to (the) territory], satisfying valency requirement of the adjective, has high preference, remaining four dependency pairs are assigned with the same, lower degree of 'reliability' of free modifiers.

B. For Pg analyzed as *na*+Loc the analysis is analogous. No valency requirement is detected, five dependency pairs are obtained based on WOP, *omezený na území* [limited on (the) territory], *způsobit na území* [caused on (the) territory], *skluz na území* [delay on (the) territory], *realizace na území* [realization on (the) territory] and *autonomie na území* [autonomy on (the) territory]. They have the same preference of free modifiers.

The possibility of evaluation of syntactic structures obtained is very important from the point of view of our task – the possibility of the detection of Pg ambiguity. It allows not to rely on the completeness of input information that, in general, cannot be granted within any automatic parsing system.

VI.   TESTING THE CRITERIA PROPOSED

In the sixth part I illustrate the possibility to incorporate the refined linguistic criteria into a concrete automatic procedure. The basis of this procedure is formed by the grammar of Robust parser for Czech (Kuboň (1999)) and its interpret (Holan (2001)). This grammar treats Pg's uniformly as free modifiers (this decision is justified by the main goal of the robust parser, i.e. the capability to cover not only well-formed but also ill-formed sentences). Thus the main changes concern the possibility to analyze Pg also as a valency or quasi-valency modifier of verbs, nouns and adjectives.

The analyses obtained are automatically evaluated according to the type of criteria used, the doubled valency/free analyses are excluded. The formal criteria and the possibility to relax word order constraints are checked manually as well the existence of verbonominal collocation in the sentence tested.

CONCLUSIONS

The present thesis concentrates on the possibility of automatic processing of prepositional groups, i.e. on the possibility to detect adequate governors of a particular Pg in the sentence tested.

I have discussed linguistic phenomena that are understood as sources of the ambiguity of Pg's. Such phenomena, 'responsible' for Pg-ambiguity, belong both to morphology and to

syntax, and consequently ambiguity of Pg's can be sorted into two types, **morphemic (structural) ambiguity** and **phonological ambiguity**.

The possibility of searching for rather complicated structures in the **Prague Dependency Treebank** has been illustrated, namely structures containing a prepositional group in such a position in a sentence that the suspicion exists that there is a morphemic ambiguity of a Pg in the sentence.

I have introduced basic principles of the **analysis by reduction**, which can naturally underlie human as well as automatic processing of natural language. The mechanism of such analysis is exemplified, application of these principles in automatic procedure for processing of Pg's is discussed.

Four types of **linguistically-based criteria** are investigated with the help of RA: criteria based on word order constraints, on valency requirements of various parts of speech, rules exploiting semantic features of single words; in addition also some formal criteria are stated.

Let us conclude that the surface position of a prepositional group is crucial for its potential ambiguity. For the detection of an adjectival governor of a particular Pg the surface word order seems to be even more important than valency requirements of the adjective. On the other hand, there are no constraints on the position of a verbal modifier expressed by a Pg, thus word order plays no role in this case. The valency requirement of a particular noun for a Pg-modifier is relevant despite the word order, if such requirement is not 'blocked' by the same requirement of a verb. The advantage of rules based on semantic features is disputable, they can be used only as supplementary criteria, if at all. On the contrary, the formal criteria – concerning complex word order restrictions as the separation principle – are of a high reliability.

The model of analysis by reduction is realized as a sequence of steps in which these criteria are applied. The criteria proposed are **arranged** and the possible analyses of prepositional group are **evaluated** according to the type of criteria leading to the establishing particular dependency pairs.

In the last part I briefly illustrate the possibility to incorporate the fine linguistic criteria to concrete parsing procedure.

I have shown that – if we exceed the field of relatively simple analysis based largely on syntactic rules – every refinement of a grammar means very fine and expensive linguistic observations far exceeding the frame of pure syntax.

The work presented is thematically limited to the ambiguity of prepositional groups and to the possibilities of their automatic attachment. Nevertheless, it can serve as a methodological basis for an automatic procedure using fine linguistic observations and covering at least 'core' of Czech.

The wide range of linguistic observations on data obtained can be used in other applicational as well as theoretical tasks – let me mention at least the building of rich Czech lexicon, the inquiry of valency frames of particular parts of speech as well as word collocations or the stating constraints on non-projective structures.

REFERENCES (selected)

J. Allen (1987)   Natural Language Understanding. The Benjamin/Cummings Publishing Company.

A. Bémová (1979) Syntaktické vlastnosti prefigovaných sloves. In: Explicite Beschreibungder Sprache und automatische Textbeargeitung V., pp. 69-194.

E. Buráňová (1980)   Ob odnoj vozmožnosti semantičeskoj klassifikacii suščestvitel'nyx. PBML 34, Prague, pp. 33-43.

Fr. Daneš (1964)   Opyt teoretičeskoj interpretacii grammatičeskoj omonimii, Voprosy jazykoznanija 6, pp. 3-16.

A. Dikovsky, L. Modina (2000)   Dependencies on the other side of the Curtain. Traitement Automatique des Langues (TAL, ed. S. Kahane), vol. 41, No. 1, Hermes, pp. 79-111.

Ch. J. Fillmore (1968)   The Case for Case. In: Universals in Linguistic Theory (eds. E. Bach, R. Harms), New York, 1-90.

Ch. J. Fillmore (1977)   The Case for Case Reopened. In: Syntax and Semantics 8 (eds. P. Cole, J.M. Sadock), New York-San Francisco-London, pp. 59-81.

J. Hajič (1998)   Building a Syntactically Annotated Corpus: The Prague Dependency Treebank. In: Issues of Valency and Meaning. Studies in Honour of Jarmila Panevová (ed. E. Hajičová), Karolinum, Charles University Press, Prague, pp. 106-132.

E. Hajičová, J. Panevová, P. Sgall (ve spolupráci s A. Böhmovou, M. Ceplovou a V. Řezníčkovou) (2000)   A Manual for Tectogrammatical Tagging of the Prague Dependency Treebank. ÚFAL/CKL Technical Report TR-2000-09.

T. Holan (2001)   Nástroje pro vývoj závislostních analyzátorů přirozených jazyků s volným slovosledem. Disertační práce, MFF UK, Praha. (manuscript)

T. Holan, V. Kuboň, K. Oliva, M. Plátek (2000)   On Complexity of Word Order. In: Les grammaires de dépendance - Traitement automatique des langues (TAL), Vol. 41, No. 1 (special issue on dependency grammars, guest editor S. Kahane), pp. 273-300.

V. Kuboň (1999)   A Robust Parser for Czech. ÚFAL Technical Report TR-1999-06.

J. Kuryłowicz (1936)   Dérivation lexicale et dérivation syntaxique. Bulletin de la Société linguistique de Paris 37, pp. 79-92.

S. Marcus (1965)   Sur la notion de projectivité. In: Zeitschr. f. math. Logik und Grundlagen d. Math. XI, pp. 181-192.

I. Mel'čuk (1988)   Dependency Syntax: Theory and Practice. NY:SUNY Press, Albany.

M. Novotný (1988)   S algebrou od jazyka ke gramatice a zpět. Academia, Praha.

L. Nebeský (1972)   A Projectivity Theorem. Prague Studies in Mathematical Linguistics 3, (Academia, Prague), pp.165-169.

K. Oliva (1996)   A Grammar Checker for Czech. JRP PECO 2824 Language Technologies for Slavic Languages, part of Final Research Report, Prague.

J. Panevová (1966)   Nesoglasovannoe opredelenie s točki zrenija analiza dlja mašinnogo perevoda. PSML 1, Academia, Prague, pp. 219-239.

J. Panevová (1974, 1975)   On Verbal Frames in Functional Generative Description. Part I, PBML 22, pp.3-40, Part II, PBML 23, pp. 17-52.

J. Panevová (1981)   K otázkám homonymie a neutralizace ve stavbě věty. Jazykovedné štúdie 16, pp. 85-89.

J. Panevová (1994)   Valency Frames and the Meaning of the Sentence. In: The Prague School of Structural and Functional Linguistics (ed. Ph. L. Luelsdorff), Amsterdam-Philadelphia, John Benjamins, 223-243.

J. Panevová (1998)   Ještě k teorii valence. Slovo a slovesnost 59, pp. 1-14.

J. Panevová (2000)   Poznámky k valenci podstatných jmen. In: Čeština - univerzália a specifika 3, Sborník z konference ve Šlapanicích u Brna (ed. Z. Hladká, P. Karlík), MU, Brno, pp. 173-180.

J. Panevová, P. Sgall: Towards an automatic parser for Czech. In: International Review of Slavic Linguistics, pp. 433-445, 1979.

J. Panevová, P. Sgall (1980)   On Some Issues of Syntactic Analysis of Czech. The Prague Bulletin of Mathematical Linguistics 34, pp. 21-32.

J. Panevová, M. Straňáková (1999)   Some Types of Syntactic Ambiguity; How to Treat them in an Automatic Procedure. In: TSD'99, Proceedings (eds. V. Matoušek, P. Mautner, J. Ocelíková, P. Sojka), LNAI vol.1692, Springer, pp. 50-55.

A. Ratnaparkhi (1998) Statistical Models for Unsupervised Prepositional Phrase Attachment. In: COLING/ACL'1998. Proceedings, pp.1079-1085.

J. J. Robinson (1969)   Case, Category and Configuration. Journal of Linguistics 6, pp. 57-80.

P. Sgall (1967)   Generativní popis jazyka a česká deklinace. Academia, Praha.

P. Sgall (1992)   Underlying Structure of Sentences and its Relations to Semantics. In: Festschrift fűr Viktor Jul'evič Rozencvejg zum 80. Geburtstag, Wiener Slawistischer Almanach 33 (ed. T. Reuther), Vienna, pp. 273-282.

P. Sgall (1998)   Revisiting the classification of the dependents; Interestingresults and tempting topics for further research. In: Issues of Valency and Meaning. Studies in Honour of Jarmila Panevová (ed. E. Hajičová), Karolinum, Charles University Press, Prague, pp. 15-26.

P. Sgall, E. Hajičová, J. Panevová (1986)   The meaning of the sentence in its semantic and pragmatic aspects. Ed. by J. Mey, Dordrecht:Reidel and Prague:Academia.

S. M. Shieber (1986)   An Introduction to Unification-Based Approaches to Grammar. CSLI Lecture Notes 4, Standford University, Stanford.

K. Sikkel, A. Nijholt (1997)   Parsing of Context-Free Languages. In: Handbook of Formal Languages (eds. G. Rosenberg, A. Salomaa), Springer, Berlin and Heidelberg, pp. 61-100.

H. Skoumalová (1994)   Czech Dictionary for the Grammar Checker. JRP PECO 2824 Language Technologies for Slavic Languages, Research Report. Saarbruecken, pp. 130-139.

H. Skoumalová (2001) Czech syntactic lexicon. PhD thesis, Charles University, Faculty of Arts, Prague. (submitted)

M. Straňáková (1998)   Ambiguity in Czech Sentences, its Classification and Searching for it. In: Proceedings of WDS'98 (ed. J. Šafránková), Matfyz-press, pp. 165-170.

M. Straňáková (1999)   Selected Types of Pg-Ambiguity. PBML 72, Praha, pp. 29-58.

M. Straňáková (2000)   Selected Types of Pg-Ambiguity; Processing Based on Analysis by Reduction. In: TSD2000, Proceedings (eds. P. Sojka, I. Kopeček, K. Pala), LNAI vol.1902, Springer, pp. 139-144.

M. Straňáková-Lopatková (2001) Ambiguity of Prepositional Groups; Classification, Criteria and Method for Automatic Processing. In: sborník university v Oldenburgu. (in press)

M. Straňáková-Lopatková (2001) Některé typy syntaktické homonymie (z hlediska možnosti automatického zpracování). In: Čeština - univerzália a specifika 3. Sborník konference ve Šlapanicích u Brna, 22.-24.11.2000 (ed. Z. Hladká, P. Karlík), MU, Brno, pp. 183-196.

Vl. Šmilauer (1966)   Novočeská skladba. SPN, Praha (2. vydání).

L. Uhlířová (1987)   Knížka o slovosledu. Academia, Praha.

L. Tesnière (1959)   Élements de syntaxe structurale. Paříž.

M. Těšitelová (1966)   O morfologické homonymii v češtině. Praha.