SELECTED TYPES OF Pg-AMBIGUITY

Markéta Straňáková

November, 99

==

Abstract

The objective of the present paper is to analyse syntactic ambiguity of prepositional groups (Pg's) within the background of dependency oriented Functional Generative Description (FGD). Several basic types of Pg-ambiguity are presented and the possibilities of searching for their occurrences in the Prague Dependency Tree Bank (PDT) is illustrated. Criteria for automatic Pg-ambiguity detection are proposed based on 1. word order patterns, 2. valency frames of verbs, nouns and adjectives (including verbonominal collocations), 3. formal criteria concerning separation principles, and 4. semantic features; the usefulness of these criteria is discussed.

I.   INTRODUCTION

The problem of syntactic ambiguity (of morphemic units in the terms of FGD, see below, section II.) belongs to the most difficult problems of any automatic procedure of syntactic parsing in any syntactic framework (be they phrase structure oriented or dependency based). The wide range of linguistic phenomena concerned can be illustrated by example (1).

Example (1):

*Japonsko a ES budou **spolupředsedat** mnohostranným **rozhovorům** o **míru** na Blízkém východě.* (PDT[1], bl103js.fs #34, shortened)

[Japan – and – EU (EC) – will – co-chair – multilateral – talks – about – peace – on – Middle East.]

[Japan and the EU will co-chair multilateral Middle East peace talks.]

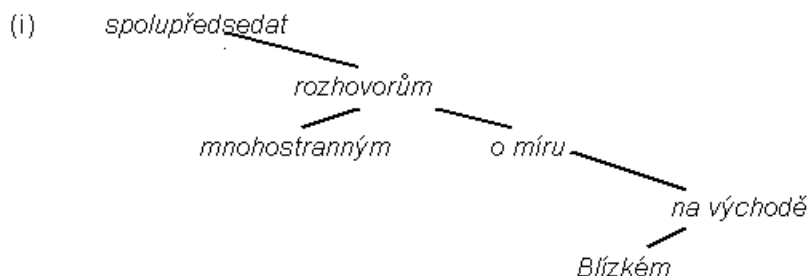The attachment of the Pg *na (Blízkém) východě* can be analysed in several ways:

Fig.1: The Pg *na (Blízkém) východě* analysed as an attributive modifier of the most recent noun *mír*
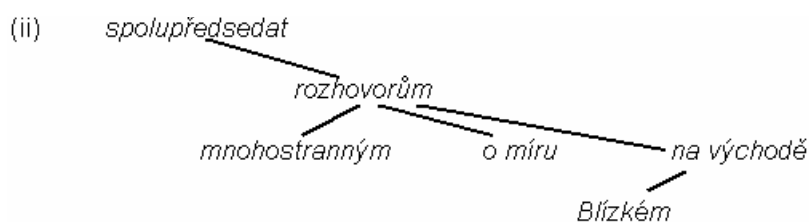
(ii)   *spolupředsedat*
       rozhovorům
mnohostranným   *o míru*   *na východě*
                           Blízkém

Fig.2: The Pg *na (Blízkém) východě* analysed as a modifier of the noun *rozhovory*

(iii)   *spolupředsedat*
        rozhovorům   *na východě*
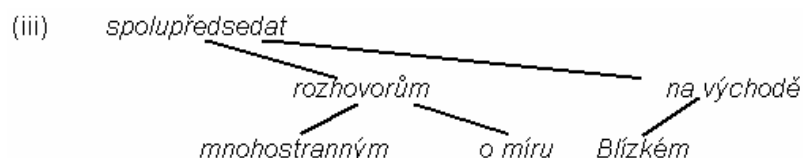*mnohostranným*   *o míru*   *Blízkém*

Fig.3: The Pg *na (Blízkém) východě* may also depend on the verb *spolupředsedat*

Syntactically we face three different structures represented by three different syntactic dependency trees. Two of them, the structures (ii) and (iii) (*spolupředsedat na (Blízkém) východě / rozhovory na (Blízkém) východě*), can be understood to denote the same situation (content) due to their implications. On the contrary, either of the pairs of structures – (i) plus (ii) (*mír na (Blízkém) východě / rozhovory na (Blízkém) východě)* on the one hand and (i) plus (iii) (*mír na (Blízkém) východě / spolupředsedat na (Blízkém) východě)* on the other – represents two different cognitive situations.

□

It is clear that the capability of the human to analyse 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, see [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, see [Kuboň, 1999] can be treated as an example of such purely syntactic approach using only limited information about particular lexical items with restricted valency frames. From our point of view the following decision is important: as it is developed to handle also ill-formed sentences, the parser processes all Pg's in the same way as free modifiers. This decision leads to a great number of cases of ambiguity of the prepositional groups (Pg's), i.e. nominal groups in prepositional cases.

The following example (2) illustrates the importance of working with a rich set of syntactic and semantic features (including complete valency frames) stored in lexicon.

[1] The possibility of searching in the Prague Dependency Tree Bank is discussed in section III.

Example (2):

Nereaguje-li nájemce <u>na nabídku</u> v průběhu 6 měsíců, ...   (PDT, bcb01aba.fs #20)

[If not respond – renter – to – offer – in – duration – (of) 6 months …]

[If the renter doesn't respond to the offer in six months …]

Fig.4: Two subtrees representing two purely syntactic analyses of Pg *na nabídku* in example (2)

The Robust Parser processes the prepositional inner participants and free modifiers in the same way. Analysing Pg *na nabídku* two syntactic subtrees are issued (Fig.4). For the purposes of a solution of Pg-ambiguity it seems, however, to be necessary to take the prepositional inner participants into account in the valency frames of verbs in the same vein as other inner participants; cf. a typical example: the verb *reagovat* needs an Object in the form *na*+Acc (Patient in underlying representation) so the object slot is filled in by the Pg *na nabídku* and the analysis in (ii) is excluded.

◻

Our goal is to investigate the possibilities of proposing the types of information that can be incorporated into an automatic procedure. We endeavour to formulate criteria which would guide either Pg-disambiguation or Pg-ambiguity detection.

Analysing sentences we take advantages of the notion of reduction analysis. Determining potential dependencies we apply the principle of deletable units, see e.g. [Jančar, Mráz, Plátek, Vogel, 1999].

The formal theory of the Robust Free-Order Dependency Grammars (RFODG) – for the definition see [Holan, Kuboň, Plátek, 1997] – serves as the formal basis of our work. The RFODG were suggested as a formal tool for developing a grammar checker for Czech. It provides the base for a parsing with subsequent localisation and evaluation of syntactic inconsistencies and errors (see the above mentioned Robust Parser for Czech, [Kuboň, 1999]).

II.   LINGUISTIC FRAMEWORK
(Definition of Ambiguity, Prepositional Group and Pg-ambiguity)

The linguistic framework of our research is formed by the dependency based Functional Generative Description (FGD) of the Czech language, see [Sgall, 1967] and [Sgall

et al., 1986]. Though the FGD was designed for the purposes of an automatic generative procedure – as its name suggests –, in fact it is 'direction independent', it can serve for an analytic procedure as well as for a generative one.

The FGD can be classed with stratificational approaches: The description of language is divided into several levels. In its 'classical' version FGD differentiates five levels – the level of underlying representation or of meaning (tectogrammatical), level of surface structure, morphemic level, morphophonemic level and phonetic level[2].

The notion of ambiguity is connected with asymmetry of the form of the sentence and its meaning. There are two ways to understand this term: (a) ambiguity (in its general meaning) as a relation between the first, phonetic level and the last, fifth level, the level of linguistic meaning, and (b) ambiguity in its more specific meaning, ambiguity on the level n, as a relation between units of two adjacent levels (i.e. levels n and n+1, where n=1,2,3,4).

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 and their common expression on the level n. In FGD terminology: two (or more) functions are expressed by the same form.

Panevová (in [Panevová, 1981]) differentiates four types of ambiguity according to the level containing the units which introduce the possibility of two (or more) readings:

Fig.5: Four types of ambiguity

Syntactic ambiguity – concerns the level of underlying representation and the level of surface structure.

---

[2] We do not discuss here the issues handled by Sgall [Sgall, 1992], i.e. those concerning the appropriateness of a level of surface syntactic structure. In the semiautomatic annotation of the corpus, characterised bellow, an intermediate level of 'analytic' structure (similar, though not identical, to surface syntax) is used as a technically conditioned means, helping to achieve a modular patterning of the complex procedure of syntactic parsing.

Morphemic ambiguity – concerns the units on the level of surface structure and their common representation on the morphemic level.

Morphophonemic ambiguity – concerns the situation when two (or more) morphemes correspond to one morphonological unit.

Phonetic ambiguity – concerns different lexical units having (accidentally) the same phonetic realisation.

□

Speaking about ambiguity in the following text we mean ambiguity on some determined level. We concentrate on the solution of a Pg-ambiguity.

From the purely syntactic point of view the morphemic ambiguity is characteristic for prepositional groups. Pg may in principal modify:

(i) any noun preceding the tested Pg in surface word order; any verb form being understood as the left-hand boundary (some exceptions must be allowed, see below section IV.2.4., verbonominal collocations);

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

(iii) any adjective.

Example (1) illustrates the types (i) and (ii), the type (iii) is shown in example (3).

Example (3):

I parlamentní demokracie musí dostát nárokům **kladeným** *na demokracii jako takovou.* (PDT, bm227zua.fs #22, shortened)

[Also – parliament – democracy – must – fulfil – claims/demands – taxed/put – on – democracy – as such.]

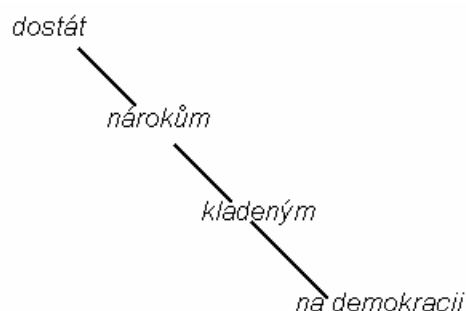[Also the parliament democracy must fulfil the demands taxed on democracy as such.]

Fig.6: The deverbal adjective *kladeným* as a governor of Pg *na demokracii*

The adjective *kladený* – derived from the verb *klást* [to tax], which constitutes a verbonominal collocation together with the noun *nárok* – is modified by Pg *na demokracii* (*jako takovou)* in this case (*klást nároky na něco* [to tax demands on something]). For the

analysis of such situation see the remarks on valency frames of adjectives bellow and also example (17) concerning verbonominal collocations.

□

Let us mention here that the fact that the adjective in example (3) is derived from a verb is not crucial – also adjectives not derived from verbs can be modified by Pg's (see example (9)).

Remark:

As some prepositions govern two (exceptionally three) morphemic cases (as e.g. *na* [on] or *v* [in], both requiring either Accusative or Locative) morphophonemic ambiguity can also appear in connection with Pg's. But whereas the morphemic ambiguity is systemic for Pg's, the occurrence of morphophonemic ambiguity is only accidental – see example (4). In this sentence both morphemic and morphophonemic Pg-ambiguity occurred:

Example (4):

*Brankář Barda a trenér Haber byli jako jediní schopni **zabezpečit přípravu** <u>na MS</u> 1993 ve Švédsku.*   (PDT, bmc26aba.fs #18, shortened)

[Goal-keeper – Barda – and – coach – Haber – were – as – the only – able – to ensure – preparation – for – World Championships – in – Sweden.]

[The goal keeper Barda and the coach Haber were the only ones who were able to ensure the preparation for the World Championships in Sweden.]

There are two types of Pg-ambiguity in this sentence.

(a) morphemic ambiguity – the Pg *na MS* can modify

   (i) either the verb *zabezpečit* (as a local modifier in Locative case)

   (ii) or the noun *příprava* (as an attributive modifier in Locative case);

(b) morphophonemic – the Pg *na MS* can modify the noun *příprava*

   (i) either as an attributive modifier in Accusative case

   (ii) or as an attributive modifier in Locative case.

It means that two distinct cases (=morphemes) are expressed by a single morphonological unit.

□

We will deal here mainly with the morphemic ambiguity of Pg's. Our task is to suggest relevant criteria – criteria usable for the purposes of an automatic procedure – allowing to detect and to describe this type of ambiguity of Pg's, i.e. to decide which word or words of a sentence can be modified by the prepositional group, if Pg-ambiguity occurs in a sentence.

In general a sentence with a Pg analysed as attributive (adnominal) modifier and the same one with the Pg analysed as verbal modifier do not have the same meaning (they usually have different truth conditions), as was illustrated in example (1). This is the reason why it is important to determine the real governor/governors of a Pg in the dependency tree representation of a sentence. The chance that their occurrences in specific contexts share their cognitive content (truth conditions) cannot be predicted in advance, it can be determined (and usually is determined) only on the basis of knowledge based inferencing.

III.   DATA GATHERING

In the first phase a sufficient amount of sentences with potentially ambiguous prepositional groups were looked for and the Prague Dependency Tree Bank (PDT) was used as the basic source.

The PDT is a corpus of Czech with rich annotation scheme, see e.g. [Hajič, 1998], which has a three-level structure: full morphological annotation on the lowest level, syntactic annotation using dependency syntax on the intermediate (analytic) level and the annotation on the level of linguistic meaning. The PDT data contain general newspaper articles, economic news and analyses and popular science magazine and information technology texts.

The structure of a sentence is represented by a dependency syntactic tree on the analytic level, the number of nodes of this tree is equal to the number of words in the sentence plus an extra root node (with some minor exceptions). The nodes are annotated by complex symbols (attribute-value pairs), with the analytical function (which reflects the relation between the dependent node and its governor) being the most important attribute. The prepositional group is represented as a subtree rooted by the preposition.

Syntactic annotation of the PDT is used for searching for pre-defined 'suspicious structures', i.e. structures which can signalise an occurrence of Pg-ambiguity. The sentences contained in the PDT are disambiguated, the only appropriate syntactic structure has been chosen manually (with respect to the context). But there are nodes in the tree representation of the sentence, which could depend also on some other governor – without any truth-conditional or situational difference between the two (or more) cases. They are marked by 'doubled' analytical functions. Sentences containing such 'double-marked' nodes have served as a basis for the definition of 'suspicious (syntactic) structures'.

Definition of 'suspicious structures':
Three basic 'suspicious (syntactic) structures' were defined:

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

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

• syntactic subtree where a verb or a noun is modified by a Pg; another noun (or another Pg')
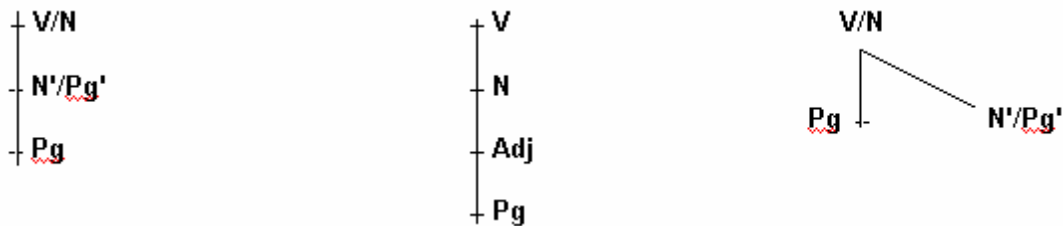appears as a brother of this Pg (i.e. (V/N (Pg N'/Pg')) in the linearised form).

Fig.7: Three basic types of 'suspicious (syntactic) structures'

□

Two macros for Graph editor (i.e. editor used for the purposes of PDT) were designed
which make it possible to search for sentences whose syntactic structure contains defined
subtrees – A-type covering the first and the second structure and B-type dealing with the third
structure. Word order is not reflected by the macros – we will return to this issue below.

At the first stage of our inquiry we focussed on Pg's with prepositions *na* [on] and *v*
[in] (both with Accusative and Locative). A sample of 1000 sentences from PDT was tested
and about 150 sentences with 'suspicious structures' were received for each preposition. The
examples (5) and (6) illustrate the results of the searching procedures.

Example (5): (B-structure, prep *v/ve*)

***Cenu** J. Seiferta, …, **zřídila Nadace** Charty 77 <u>ve Stockholmu</u> <u>v lednu</u> 1986 k **uctění** památky
prvního čs. nositele Nobelovy ceny za literaturu.* (PDT, bmb09eba.fs   #19, shortened)

[Award of J. Seifert – , …, – established – Charta 77 Foundation – in – Stockholm – in –
January – 1986 – to – honour – (of) memory – (of) first – Czechoslovakian – bearer – (of)
Nobel Prize – for – literature]

[J. Seifert Award, …, was established by the Charta 77 Foundation in Stockholm in January
1986 in honour of the memory of the first Czechoslovakian bearer of Nobel Prize for
literature.]

The Pg's *ve Stockholmu* and *v lednu 1986* modify the verb *zřídit* in the PDT, but they
can be also treated as an attribute of the noun *Nadace*. As the word order conditions are not
reflected, the macro proposes also the nouns *cena* and *uctění* as potential governors of these
Pg's.

□

Example (6): (A-structure, prep *na*)

*V přímém přenosu **přivítají** například i **půlnoc** <u>na Staroměstském náměstí.</u>*

<div align="right">(PDT, bl110js.fs #18)</div>

[In – live – broadcast – (they will) welcome – e.g. – also – midnight – at – Old Town Square]

[They'll welcome e.g. also midnight at the Old Town Square in live broadcast.]
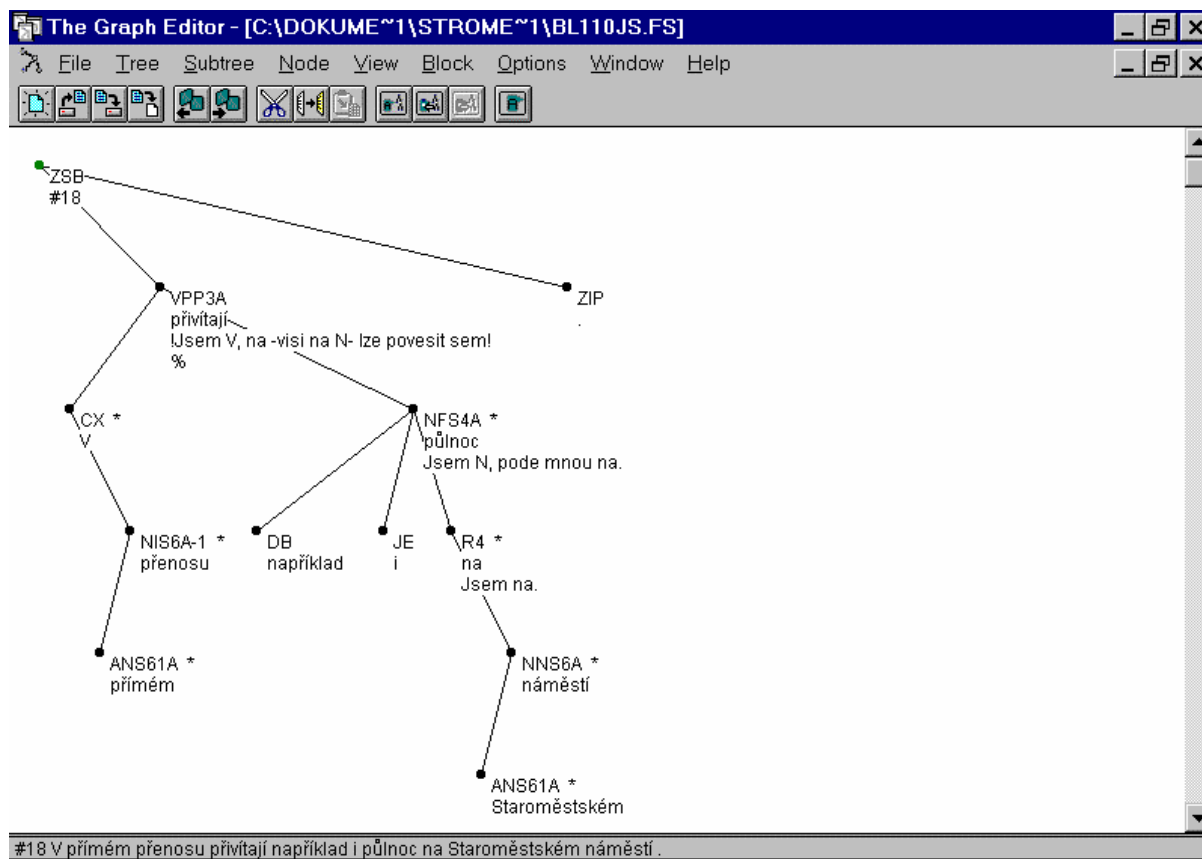

Fig.8: The result of the searching procedure for preposition *na* in the sentence (6)

The Pg *na (Staroměstském) náměstí* is treated as an attributive modifier of the noun *půlnoc* (afun Atr) in PDT, but it be can also treated as a local modifier of the verb *přivítat* (afun Adv).

□

As mentioned above, word order is not reflected by the macros as was mentioned above. This decision causes that – searching for the occurrences of potential Pg-ambiguity – also inappropriate 'syntagms' are found by the macros (as in example (6)). But on the other hand it brings us one important advantage. We want to investigate also Pg-modifiers extending verbonominal collocations that are treated as non-congruent attributes. In these cases the word order limits (as they are described in the section IV.1.) can be omitted by Pg-

modifiers (for example, such Pg's may lead to non-projective constructions, see section IV.2.4.). So by ignoring the word order restrictions the macros can capture also all occurrences of the Pg-modifiers extending verbonominal collocations.

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

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

• B-type macro examines all brothers of the node representing the tested Pg;

• both A-type and B-type macros handle the coordination in any position of the defined schemas, so sentences like the one in example (7) are also chosen.

Example (7): (coordination, prep *na*)

Stát neposkytne <u>*na nákup*</u> *bytů žádné* **slevy***, ani žádnou jinou finanční* **pomoc***. (PDT, bcb01aba.fs #16, shortened)

[State – will not provide – for – purchase – (of) flats – no – discounts – nor – any – other – financial – help.]

[The state will not provide any discounts for apartment purchase or any other form of financial help.]

The Pg *na nákup* is treated as an adverbial modifier (Adv) of the verb *neposkytnout* in the PDT; the B-type macro offers also the nouns *slevy* and *(finanční) pomoc* as the potential governors of the tested Pg's. The noun *stát* is also offered, as word order restrictions are not considered.

In fact, the verb *poskytnout* and the noun *sleva* constitute a verbonominal collocation *poskytnout slevu na něco* [to provide discounts for something], the Pg *na nákup* can be understood as their 'common' modifier (see section IV.2.4.).

□

IV.   PROPOSAL OF THE CRITERIA

The appropriate disambiguated syntactic analysis of a sentence (and especially of its parts dealing with Pg's) is usually based not only on purely syntactic rules but also on the semantic and even pragmatic knowledge available to human beings. These semantic and pragmatic aspects are too complex to be included in an automatic procedure in their entirety. We try to detect the types of input information the parser can take into account and to determine the criteria both for the detection of Pg-ambiguity or for Pg-disambiguation.

Here we present the prerequisites necessary for the setting up of the criteria and for their subsequent application.

Lexicon:

• The existence of lexicon with detailed information is presupposed. The lexicon must contain complete valency frames for verbs, nouns and adjectives – both the 'first order' and the 'second order' valency slots (see below, section IV.2.) are to be included.

• The verbonominal collocations must be kept in the lexicon with their valencies.

• The semantic features of nouns and verbs are also stored in the lexicon together with possible semantic functions of prepositions.

Pre-processing:

• The data provided by full morphological analysis is used as an input information the criteria operate on.

• Some additional (mainly local) information is presupposed: an identification and analysis of analytical verb forms, connection of noun groups with numerals, identification of frozen collocations and verbonominal collocations and also some estimation of boundaries of clauses.

IV.1. Word Order Patterns

The surface word order criteria serve as the basic purely syntactic clue to the Pg-ambiguity detection. Four observations lead to the specification of the basic word order patterns with respect to Pg-disambiguation:

Observation (1):

The prepositional group as a noun modifier prototypically modifies only nouns preceding this Pg in the surface representation of the sentence; i.e. the governor of an attributive Pg modifier must in principle precede this Pg.

This observation is not valid for verbonominal collocations – the nominal part of such unit (which can be treated as a governor of examined the Pg) can stand after this Pg modifier constituting a non-projective construction (see below, section IV.2.).

□

Observation (2):

Any verb form serves as a block for Pg attachment in the primary case – the prepositional group cannot modify any noun separated from it by any verb form, i.e. no verb form can appear between Pg treated as an attributive modifier and its noun governor in the surface representation of a sentence. (This is not valid for specific cases, esp. For verbonominal

collocations, where the Pg modifying the noun part stands at the very beginning of a clause, creating non-projective construction.)

□

Observation (3):

Any verb form serves as a block also for the attachment of a Pg as a modifier of an adjective – the prepositional group cannot modify an adjective separated from it by any verb form in the surface level.

□

Observation (4):

The Pg modifying an adjective can either precede or follow its governor in the surface representation of a sentence.

□

Before defining the basic unambiguous and ambiguous word order patterns with respect to Pg-disambiguation, we have to describe here the notion of analysis by reduction, see e.g. [Jančar, Mráz, Plátek, Vogel, 1999].

Description of analysis by reduction:

The analysis by reduction consists of stepwise simplification of an extended sentence so that its syntactic correctness is preserved. In each step the simplification is realised by deleting one word of a sentence and possibly rewriting other words. This process is non-deterministic, in each phase any of mutually independent words can be deleted[3].

□

This basic approach to the analysis of a sentence has crucial influence on the shape of the word order patterns – all rules are applied repeatedly in different stages of analysis. (We will return to this point later, see section V. about the arrangement of criteria, especially the continuation of the remarks on example (8)). Inquiring the possibility of nominal or verbal governors of Pg we take into account such branches of the reduction process where the potential congruent adjectives have been deleted (they are treated as congruent attributes of nouns). Therefore, they do not appear in the word order patterns. However, if we search for a potential adjective governor of Pg, those branches of reduction analysis are chosen in which adjectives are still preserved and thus expressed in word order patterns. (Generally all

---

[3] We will return to the notion of analysis by reduction in the section V. where also several examples are introduced to show the mechanism of such analysis.

dependent words must be deleted before their governor is reduced, otherwise the condition of correctness preservation during analyses by reduction is not fulfilled.)

Let us now introduce the basic word order patterns with respect to Pg-disambiguation based on the four above mentioned observations (1)-(4):

Definition of unambiguous word order patterns:

Three basic unambiguous word order patterns with respect to Pg-disambiguation (concerning the surface representation of a sentence) are defined:

• V (resp. N at the beginning of a clause) immediately followed by Pg   (such V, resp. N, is the Pg governor);

• Pg at the very beginning of a clause (V is the Pg governor in this case – the only exception of verbonominal collocations is taken into account – see below);

• Adj and Pg are separated by any verb form (V is the Pg governor in such case)

(e.g.   Pg – V – Adj – N   and   Adj – N – V – Pg).

□

Definition of 'suspicious' word order patterns:

Three basic 'suspicious' word order patterns with respect to Pg-disambiguation (concerning the surface representation of a sentence) are specified:

• Pg between a noun and a verb (from the left to the right) (i.e.   N – Pg – V);

• the sequence of a verb, a noun and Pg (in this order) (i.e.   V – N – Pg);

• Pg between a verb and an adjective or Pg being followed by an adjective and a verb (in this order)

(i.e. V – Pg – Adj – N ,   Pg – Adj – N – V ,   Pg – Adj – V – N ,   and also   Adj – N – Pg – V).

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.

□

Observation (5):

A congruent adjective (without its own modifiers) usually precedes its governing noun in Czech (with exception of specific contrastive positions and of scientific terminology as e.g. in biology or chemistry). Thus the fact that the adjective follows its noun governor signalises that such adjective is modified, often by the subsequent Pg.

□

The word order patterns concerning the occurrence of an adjective can be illustrated by the examples (8) and (9).

Example (8):

*Dívka, ..., rovná <u>na ramínku</u> **vystavený** kabát.* ([Panevová,1998a])

[Girl – … – arranges – on – hanger – exposed – coat.]

[A girl … arranges an exposed coat on a hanger. / A girl … arranges a coat exposed on a hanger.]

The Pg *na ramínku* may depend either on the preceding verb *rovnat* or on the following deverbal adjective *vystavený* being local modifier in both cases.

Further 'suspicious' word order variations:

(i) <u>*Na ramínku*</u> **vystavený** *kabát* **rovná**.

(ii) <u>*Na ramínku*</u> **vystavený** **rovná** *kabát*.

(iii) **Vystavený** *kabát* <u>*na ramínku*</u> **rovná**.

The first and the second variants are undoubtedly ambiguous with no doubts (if we could admit non-projective construction *vystavený rovná kabát* in (ii)). In the third variant the non-projective construction **vystavený** *kabát* <u>*na ramínku*</u> is not stylistically recommended (see [Uhlířová, 1987]) but it is commonly used.

Unambiguous word order variations:

(i) *Vystavený kabát* **rovná** <u>*na ramínku*</u>.

(ii) <u>*Na ramínku*</u> **rovná** *vystavený kabát*.

(iii) *Kabát* **vystavený** <u>*na ramínku*</u> *rovná*.

(iv) *Rovná kabát* **vystavený** <u>*na ramínku*</u>.

The verb *rovnat* separates Adj and Pg in the first and the second variant – the Pg *na ramínku* cannot modify the adjective. In the variants (iii) and (iv) the deverbal adjective *vystavený* follows its noun governor and so – according to the observation (5) – the subsequent Pg is treated as its modifier.

(For stepwise reduction analysis of original sentence see continuation in section V.)

□

The following example (9) shows the Pg-ambiguity concerning the verb and an adjective not derived from a verb; Pg is treated as a free modifier in this sentence.

Example (9):

<u>*Na první pohled*</u> **hrůzostrašný** *sjezd* **není** *nebezpečnější než jiné lyžařské disciplíny.* ([Kuboň, 1999])

[On – first – sight – horrific – downhill run – is not – more dangerous – than – other – ski – events.]

[The downhill run, which is horrific on the first sight, is not more dangerous than other ski events. / On the first sight, the horrific downhill run is not more dangerous than other ski events.]

The Pg *na (první) pohled* can be analysed in two ways: either as a modifier belonging to the adjective *hrůzostrašný* or as a free verbal modifier.

☐

Special interest must be devoted to the chains of nouns where (at least one) noun in the genitive case appears followed by a Pg. Such sentences seem to be 'very hot candidates' for the occurrence of Pg-ambiguity – the possibility of two analyses mentioned in example (10) is systemic in Czech.

Example (10):

*V Gale nepředpokládají, že by se **trh práce** <u>na Prostějovsku</u> do dvou let zásadně **změnil**.*
(PDT, bcd11eba.fs #48, shortened)
[In – Gala – not-expect – that – would – market – (of) labour$_{gen}$ – in Prostějov region – to – two – years – markedly – change.]
[At Gala they do not expect / Gala does not expect the labour market in the Prostějov region to change markedly within two years.]

The Pg *na Prostějovsku* can be treated as an attributive modifier of either of two nouns *trh* and *práce* (the noun *práce* is a non-congruent attribute in Genitive case of the noun *trh*). (In addition the Pg *na Prostějovsku* can also depend on the verb *změnil*.)

☐

IV.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 notions of valency, valency frames and valency slots are crucial for the underlying representation of a sentence, see e.g. [Panevová, 1974, 1975], [Panevová, 1994]. For the purposes of syntactic analysis of natural language sentences the valency theory has impact also on the level of surface representation.

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

IV.2.1. Valency Frames on the Level of Underlying Representation

Originally the valency theory as a substantial part of the FGD was established for verbs and their frames, see [Panevová, 1980], the extension on nouns and adjectives followed, [Piťha, 1981], [Piťha, 1982] and [Panevová, 1998].

Verbal Valency Frames

The valency frame of particular lexical item (in the narrower sense of 'frame')[4] consists of the inner participants (i.e. Actor, Patient, Addressee, Origin and Effect), either obligatory or optional – with the list of one or more appropriate morphemic forms – and of the obligatory free modifiers (as local, temporal, manner, casual etc). Marked obligatory members of a valency frame can be omitted on the surface level. Each Czech verb, depending on the number of its valency frames, is represented by one or more lexical items in the lexicon.

The role of verbal valency frames in Pg-disambiguation was illustrated in example (2).

Valency Frames of Nouns

There is a slightly more complicated situation concerning valency frames of nouns since the nouns derived from verbs must be distinguished from the other nouns.

The **deverbal nouns**, i.e. nouns derived from verbs, inherit the frames from the original verbs, so that they have the same repertoire of inner participants. Some regular changes are met in the surface representation, e.g. Nominative case is regularly transformed into Genitive case, Accusative case into Genitive case etc. None of the valency frame members is obligatorily present on the surface. Example (11) shows the changes in the frame of deverbal noun and its application for Pg-disambiguation.

Example (11):

*K **nominaci** Jiřího V. Kotase na čs. prezidenta předseda Strany zelených Aleš Mucha poznamenal, že ...*   (PDT, bm122zua.fs #2)

[To nomination – (of) Jiří V. Kotas – for – Cs. – president – chairman – (of) Green Party – Aleš Mucha – noticed – that – …]

[(Commenting) on the nomination of Jiří V. Kotas for the Czech president the chairman of the Green Party Aleš Mucha noted that …]

---

[4] Most of the free modifiers are optional and belong only to a 'valency frame' in a broader sense.

The noun *nominace* is derived from the verb *nominovat* [to nominate], which has three valency slots in its frame – for Actor (in Nominative), for Addressee (in Accusative), and for Patient (with the surface form *na*+Acc). The valency frame of the noun *nominace* is preserved with three possible expressions – (i) Actor (possessive adjective), Addressee (in Genitive), Patient (*na*+Acc), (ii) ambiguous possessive adjective (corresponding either to Actor or to Addressee), Patient (*na*+Acc) or (iii) Addressee (possessive adjective), Patient (*na*+Acc), possibly with Actor in Instrumental.

nominovat … Act / Adr Pat

někdo$_{Nom}$ *nominuje někoho*$_{Acc}$ *na něco*$_{Acc}$ → *něčí*$_{subj}$ *nominace někoho*$_{Gen}$ *na něco*$_{Acc}$

→ *něčí*$_{sub/obj}$ *nominace na něco*$_{Acc}$

→ *něčí*$_{obj}$ *nominace na něco*$_{Acc}$ *někým*$_{Ins}$

As the Pg *na (čs.) prezidenta* fills in the requirements for Pat modifier of the noun *nominace* it is analysed as depending on this noun. (The analyses as *Kotase na (čs.) prezidenta* or *poznamenal na (čs.) prezidenta* are excluded.)

□

Other nouns (i.e. nouns not derived from verbs) have the same repertoire of complementations as the verb plus the set of specific adnominal modifiers (i.e. Partitive or Material, Appurtenance, Identity, Restrictive and Descriptive Adjunct; Partitive and Identity are considered to be inner participants); see [Sgall, Hajičová, Panevová, 1986] and example (12). The valency frames of nouns must be stored in the lexicon.

Example (12):

šance na byt   [chance for flat]

šance … Pat (*na*+Acc)

právo na uveřejnění   [right on publication]

právo … Pat (*na*+Acc)

zákon na ochranu   [law for protection]

zákon … Aim (*na*+Acc)

názor na další vývoj   [opinion on further development]

názor … Pat (*na*+Acc)

slevy na nákup   [discounts on purchase]

sleva … Pat (*na*+Acc)

□

Valency Frames of Adjectives

The valency frames of adjectives have been studied by Piťha and Panevová, see [Piťha, 1982] and [Panevová, 1998]. They start from **deverbal adjectives** (i.e. adjectives derived

from verbs) which have the same repertoire of inner participants as verbs. A deverbal adjective shares its valency frame with the original verb. Two regular differences are present:

• None of the valency frame members is obligatory, each of them can be omitted.

• One of the presupposed valency slots is filled in by the word that is modified by the examined adjective (i.e. by the governor of the adjective, see [Panevová, 1998b]).

Let us return to example (3) to illustrate the relation between the verb and the adjective derived from it:

Example (3): (continuation)

klást   [to tax / to put]

někdo$_{Nom}$ *klade něco*$_{Acc}$ *někam / na něco*$_{Acc}$ →   *něco kladené* (*někým*$_{Ins}$) (*někam / na něco*$_{Acc}$)

(*něčí*) *něco kladené* (*někam / na něco*$_{Acc}$)

[somebody – puts – something – somewhere / on something]

[something – put – (by somebody) – somewhere / on something]

[(somebody's) – something – put – somewhere / on something]

klást   …   Act / Pat Loc   →   *kladený*   …   Act / Pat Loc

Compare also with example (17).

□

In addition to the same list of inner participants and free modifiers as the verbs have, **non-deverbal adjectives** have also modifiers of the second and the third degree. Their valency frames must be stored in the lexicon, too.

IV.2.2. Valency Frames on the Level of Surface Representation

The interpretation of the valency frames (defined on the level of underlying representation) on the surface level is suitable from the syntactic parsing point of view. Some regular changes are to be done – first of all, each member of the valency frame of the particular verb, noun or adjective can be omitted on the surface level, so the possibility of the omission must be extended to all inner participants as well as to all free modifiers.

On the other hand some rules seem to be valid for the omission of valency frame members on the surface level.

Observation (6):

If the optional inner participant Addressee, Origin or Effect of a particular verb is expressed on the surface level, then an obligatory Patient must be expressed there, too. This observation is valid for the verbs with participants in 'prototypical forms' as in the next example (13).

Example (13):

Ušila dětem hračku ze zbytků látky. (Panevová)

[(She) sewed – (to) children – toy – from – remains – (of) cloth.]

[She sewed a toy for children from the remains of cloth.]

The valency frame of the verb *ušít* consists of Actor, obligatory Patient (*hračka*), optional Addressee (*děti*) and optional Origin (the Pg *ze zbytků (látky)* is treated as a verbal modifier here):

ušít … Actor / Patient (Addressee) (Origin)

The presence of Addressee (*děti*) or Origin (*ze zbytků (látky)*) on the surface level leads to the necessity of expressing the Patient (*hračka*) there – compare the correctness of the following modifications of the original sentence:

Ušila hračku ze zbytků látky.

Ušila dětem hračku.

*Ušila dětem ze zbytků látky.

Ušila hračku.

*Ušila dětem.

*Ušila ze zbytků látky.

While the first, the second and the fourth sentence – where Addressee, Origin and the both, Addressee and Origin, respectively, are omitted – are quite well-formed, the sentences with the omitted Patient are ill-formed.

□

Such regularities in the deletion of valency frame members on the surface level of representation require a further linguistic research.

IV.2.3. The 'Second Order' Valency Frames

Panevová in [Panevová, 1966] proposes the 'second order' valency describing (nominal) modifiers which cannot be treated as valency modifiers in (strictly) grammatical sense for the purposes of syntactic analysis. The advantages of such 'second order' valency frame are very well applicable for Pg-disambiguation. As it describes commonly used modifications of the lexical items, the 'second order' valency concerns verbs, nouns as well as adjectives. Similarly as the real valency slots, the slots asking for 'common modifications' – together with their surface expression (usually several alternatives) – must be contained in the valency frame of the particular lexical item.

The **idiomatic usage of the verb** is typically described by the 'second order' slots.

Example (14):

*A Slováci <u>na tento fakt</u> **hřeší**.* (PDT, blc03zu.fs)

[And – (the) Slovaks – of – this – fact – take (undue) advantage.]

[And the Slovaks take (undue) advantage of this fact.]

The verb *hřešit* [sin / take undue advantage of st / offend] can be used in three ways:

(i) *někdo*_{Nom} *hřeší₁ (něčím*_{Ins}*) (proti něčemu*_{Dat}*)*

[somebody – sins – (by something) – (against something)]

    hřešit … Act / (Mean) (Benefactor)

i.e. with Actor and two optional modifiers, one in Instrumental case (Mean), the other in the form *proti*+Dat (Benefactor);

(ii) *někdo*_{Nom} *hřeší₂ na něco* [somebody – takes undue advantage – of something] idiomatic usage

    hřešit … Act / Pat

i.e. with Actor (in Nominative) and Patient expressed by the Pg *na*+Acc;

(iii) *někdo*_{Nom} *hřeší₃ na někom* [somebody – offend – against somebody] idiomatic usage

    hřešit … Act / Pat

i.e. with Actor and Patient expressed by *na*+Loc.

The first usage is captured in the 'first order' valency frame, other two possibilities are described by the 'second order' valency frames.

In this sentence *hřešit₂* is chosen due to its valency characteristics.

□

Several types of the **'second order' nominal modifiers** (i.e. the attributes which cannot be treated as valency modifiers in strictly grammatical sense though the noun and such attribute form the only denomination unit, see [Mathesius, 1942]) must be distinguished:

(i) The scientific and expert terminology as well as usual collocations commonly used in journalistic texts is a good example of nouns and their 'second order' modifiers. The terminology of the domains concerned must be contained in the lexicon.

(ii) Deverbal nouns can have the 'second order' modifiers derived from the obligatory free modifiers of the original verbs. Such modifiers are to be listed in the valency frames of deverbatives.

(iii) The 'second order' valency may have its origin in the separation of the parts of verbonominal collocation (see below): such collocation is used where the verb states the relation between the noun and its modifier (often expressed by Pg); in the following sentences the verb can be omitted and the modifier is treated as the 'second order' modification of the noun. The solution of this problem is closely connected with the treatment of deletions.

Example (15) illustrates the noun derivation and the changes in the valency frame (point (ii)):

Example (15):

*Nechtěný **pobyt** <u>na čerstvém vzduchu</u> nabízí svým žákům škola ... .*

<div align="right">(PDT, bm102zua.fs #15, first part)</div>

[Involuntary – stay – on – fresh – air – offers – (to) its – pupils – school … .]

[The involuntary stay on the fresh air is offered by the school to its pupils … .]

The verb *pobývat* requires an obligatory local modifier, which is preserved as 'usual' local modifier of the deverbal noun *pobyt*. The analysis of the Pg *na čerstvém vzduchu* as a local attribute of the noun *pobyt* is preferred.

□

IV.2.4. Verbonominal Collocations

The identification of the occurrences of verbonominal collocations seems to be a very effective tool for Pg-disambiguation. A verbonominal collocation is a collocation of noun and verb (often with very general meaning), the noun part has usually valency slot (either of the 'first order' or of the 'second order') filled in by the Pg.

Example (16):

mít / potřebovat / vynaložit peníze na něco   [to have / need / spend money on something]

mít / poskytnout / získat prostředky na něco   [to have / provide / gain means for something]

podat odvolání proti něčemu   [to submit appeal against something]

klást nároky na něco   [to tax demands on something]

□

Two types of verbonominal collocations can be distinguished:

• verbonominal collocation consists of a verb with very general meaning and of a noun with valency requirements;

• both the noun and the verb constituting verbonominal collocation have valency slots requiring the examined Pg – for the consequences see especially examples (17) and (18).

From the purely syntactic point of view there are two systematic possibilities how to analyse such collocations – the Pg can be treated as:

(i) an attributive nominal modifier (the observations (1) and (2) concerning word order restrictions (see above) are not valid for verbonominal collocations – the noun part of such unit can stand after its Pg modifier and they can be 'separated' by verb form);

(ii) a verbal modifier, which either is registered in its (narrower) valency frame, or is an optional free modifier.

These two structures usually share the same content (they have the same truth conditions) though they have two representations both on the surface level and on the level of underlying representation. However, also verbonominal collocations exist where both noun and verb can have the same valency bound Pg modifier – in such cases the sentence can denote two different situations depending on the choice of the governor of Pg.

Solution:

Concerning Pg-disambiguation two issues must be solved:

(a) If <u>both the name and the verb</u> constituting the verbonominal <u>collocation have a valency slot requiring the examined Pg</u> then two structures must be preserved – the one treating Pg as an adnominal modifier and the other with verbal Pg modifier.

(b) Else it seems to be sufficient to save only one of these parallel structures (the other one can be added whenever it is necessary) – the one treating Pg as an adnominal modifier in agreement with the valency requirements of the noun.

The verbonominal collocations must be kept in the lexicon among the entries of the noun item.

□

The following examples (17) and (18) illustrate consequences of the proposed solution.

Example (17):

***Kladl** (vysoké) **nároky** <u>na demokracii / na bezpečnost / na Petra</u>.* (based on PDT, bm227zua.fs #22, see ex. (3) above)

[(He) taxed/put – high – claims/demands – on – democracy / on safety / on Peter.]

[He taxed the high demands on democracy / on safety / on Peter.]

The verb *klást* constitutes a verbonominal collocation together with the noun *nároky*. There are two possible syntactic structures describing this sentence:

(i) The Pg *na demokracii / na bezpečnost / na Petra* fills in an obligatory valency slot of the verb *klást*.

někdo klade něco$_{Acc}$ *někam*

    klást … Act / Pat Loc

(ii) The noun *nároky* is modified by the Pg *na demokracii / na bezpečnost / na Petra* in agreement with its valency requirements.

    nárok … Pat (*na*+Acc)

According to the proposed solution – the point (a) – both structures must be preserved.

□

Example (18):

*Finanční prostředky <u>na nákup</u> technického vybavení a učebnic **poskytly**....* (bm122zua.fs #21)

[Financial – means/sources – for purchase – (of) technical – equipment – and – textbooks – provided – …]

[Financial sources for technical equipment and textbooks were provided by … .]

The verb *poskytnout* constitutes a verbonominal collocation together with the noun *prostředky.*

    prostředek … Aim (*na*+Acc)

The Pg *na nákup* can be treated as a nominal modifier filling in the Aim slot of the valency frame of the noun *prostředky* or it can play the role of a free verbal modifier of the verb *poskytly* (also Aim). According to the proposed solution – the point (b) – the structure with the noun chosen as the Pg governor is preserved (the other one is excluded).

    □

The fact that the Pg modifies to the whole verbonominal collocation and cannot be seen as a purely nominal modifier is reflected in the changes in syntactic structure of such groups during the adjective derivation. Let us return to example (18) to illustrate these changes concerning the derivation:

Example (18): (continuation)

poskytnout prostředky na něco   →   *prostředky poskytnuté na něco*
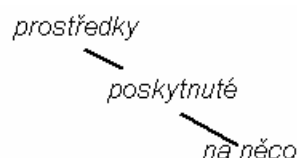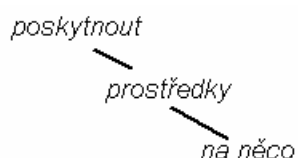[to provide – sources – for – something]   [sources – provided – for – something]

Fig.9: Deverbal adjective – the changes in syntactic structure

The structure with the nominal governor of Pg is chosen as the adequate analysis according to the proposed solution. Deriving the adjective from the verb the distinct syntactic structure is relevant: the deverbal adjective *poskytnutý* (pl *poskytnuté*) is a congruent attribute following its governing noun *prostředky*; according to observation (5) the Pg *na něco* is treated as modifier of the adjective.

    □

Non-projective Constructions

As was mentioned above, some cases exist where the observations (1) and (2) concerning word order restrictions (see above, section IV.1.) are not valid – the noun governor can follow its Pg modifier. The violation of these restrictions seems to be possible only if there is some valency requirement that the Pg fills in.

Observation (7):

• A <u>valency modifier</u> of a noun expressed by Pg can precede its noun governor in the surface structure of a sentence. Analysing such Pg modifier, we can get the non-projective structure on the level of surface representation (the violation of observation (1)).

• A <u>valency modifier</u> of a noun expressed by Pg and its noun governor can be 'separated' by the verb form (the violation of observation (2)).

• <u>Free modifiers</u> of a noun expressed by Pg's must follow their governor on the surface level; observation (2) is valid – the Pg and its noun governor cannot be 'separated' by the verb.

Example (19):

*Na poslední místo v týmu **měl** největší **šanci** některý z mladíků - Bielčík, Lukeš, Rydval, Jiroutek.* (PDT, bl123jsa.fs #20)

[On – last – position – in – team – had – (the) greatest – chance – one – of – youngsters – Bielčík – Lukeš – Rydval – Jiroutek.]

[One of the youngsters – Bielčík, Lukeš, Rydval, Jiroutek – had the greatest chance to get the last position on a team.]

The noun *šance* – constituting a verbonominal collocation together with the verb *mít* – has a valency slot asking for *na+Acc* modifier; the Pg *na (poslední) místo (v týmu)* meets this requirement creating a non-projective construction.
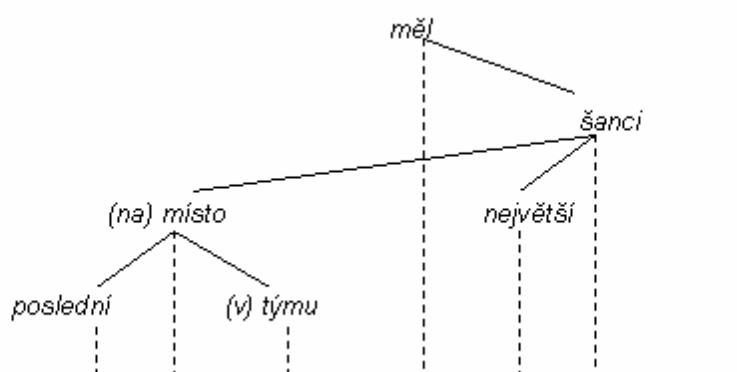

Fig.10: Non-projective construction caused by the position of the Pg valency modifier of the noun *šance*

□

The **verbonominal collocations** are the main source of this type of non-projective constructions caused by Pg modifiers of nouns. Though examples of other types of non-projective constructions with non-congruent Pg-attribute of noun are mentioned in linguistic literature (example (20) illustrates this possibility, [Šmilauer, 1947]) they do not appear in our sample of PDT and such constructions are not very frequent in texts.

Example (20):

*Od zahrady se ztratil **klíč**.* ([Šmilauer, 1966, s. 195])

[From – garden – (he) lost – key.]

[He lost a key from a garden.]

In this sentence the Pg *od zahrady* is intuitively treated as a nominal modifier preceding its governor (as the noun *klíč* has adequate 'second order' valency slot and it has no sense to analyse it as a verbal modifier) though *ztratit* and *klíč* do not constitute a verbonominal collocation.

□

Decision:

We have decided to confine to the verbonominal collocations as the only possible bearers of the non-projective constructions with non-congruent Pg-attribute of noun in our further research.

□

IV.2.5. Reliability of Valency Frames

In our discussion of verbonominal collocations we have seen that some overlappings of valency requirements may exist (see example (17)) – both noun and verb can have a modifier of the same form. Such a situation is not limited only to the verbonominal collocations, the overlapping can hit verbs as well as nouns and adjectives and their combinations.

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 of the structures satisfying valency requirements of any unit within a parsing procedure. The observation (3) concerning the dependency of Pg on adjective – any verb form serves as a block for the attachment of Pg – seems to be the only rule which cannot be violated.

Evaluation:

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

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

• The valency requirements of adjective follow (if the word order restrictions are fulfilled).

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

• The 'second order' valency frame of verb is of lower reliability.

• The 'second order' valency frame of noun has the lowest priority.

This classification of valency requirements serves to an evaluation of the final syntactic structures of a sentence with their overlapping, to classifying the structures according to their reliability.

□

The following example illustrates the situation where there are two structures fulfilling the valency requirements of different priorities.

Example (21):

Rozšíření *silniční* ***daně*** <u>*na všechna vozidla*</u> *- to je otázka, o níž podnikatelé nejčastěji hovoří.*
(PDT bcb01aba.fs #37, shortened)

[Extension – (of) road – tax – to – all – vehicles – that – is – question – about – which –
entrepreneurs – most often – speak.]

[The extension of the road tax to all vehicles – that is the question the entrepreneurs speak about most often.]

The Pg *na (všechna) vozidla* can be treated either as a modifier of the deverbal noun *rozšíření* (the 'first order' valency requirement, Effect on the level of underlying representation) or as a modifier of the noun *daně* (the 'second order' valency requirement; it is one of three alternative forms – *na*+Acc, *za*+Acc, *z*+Gen). Both structures are preserved, the first one being of a 'higher degree' of reliability.

□

Though the application of valency information is a relatively very safe criterion, some sentences were found where the valency requirements cause the creation of an inadequate syntactic structures as in examples (22) and (23):

Example (22):

Domy jsou vraceny často v žalostném stavu, ***zákon dbá*** <u>*na ochranu*</u> *nájemníků snad až příliš.*
(PDT, bl101js.fs #18, shortened)

[Houses – are – returned – often – in – terrible – state – , – law – looks – after – protection – (of) tenants – perhaps – even – too.]

[Houses are often returned in terrible state, perhaps the law looks only too much after the protection of the tenants.]

The Pg *na ochranu (nájemníků)* is analysed in two ways – either as the 'first order' modifier of the verb *dbát* (filling in the Patient slot, 'higher priority') or as the 'second order' modifier of the noun *zákon* (as *zákon* requires Aim modifier with the form *na*+Acc or *o*+Loc, 'lower priority'). But according to the observation (2) the prepositional group cannot modify any noun separated from it by a verb and so the second structure is excluded (*zákon* and *dbát* do not consist the verbonominal collocation). (For a stepwise reduction of shortened sentence, see section V.)

□

Example (23):

*V období, kdy prudce **poklesl zájem** <u>na domácím trhu</u>, dokázala továrna část výroby exportovat.*   (PDT, 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 fell sharply the factory managed to export part of its production.]

Primarily the Pg *na (domácím) trhu* is inadequately analysed as a valency modifier (Patient) of the noun *zájem*. In fact it is preferably treated as a free local modifier belonging either to the noun *zájem* or to the verb *poklesl*. (The valency slot of Pat with *zájem* is occupied by a General participant in this case.)

□

IV.3. Formal Criteria

IV.3.1. 'Separation principle'

In section IV.2. we took into account the valency information of single words of the sentence. Now we must concentrate on word order configurations again. Another important observation is to be formulated.

Observation (8):

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 analysed as a verbal participant or free modifier, 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).
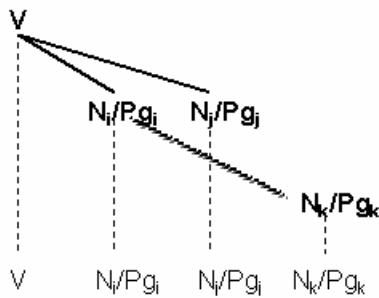

Fig.11: The impossible dependency configuration in $V - N_i/Pg_i - N_j/Pg_j - N_k/Pg_k$ chain
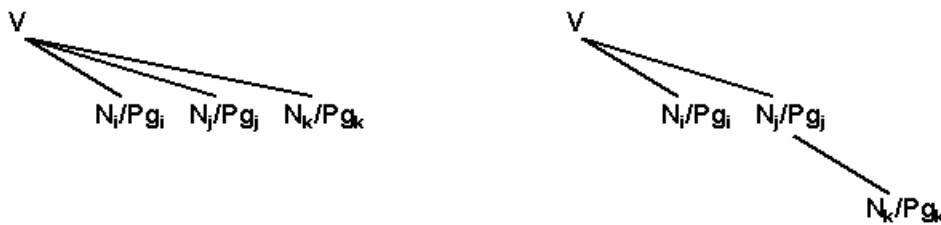

Fig.12: Two possibilities of the dependencies in $V - N_i/Pg_i - N_j/Pg_j - N_k/Pg_k$ chain

□

Example (24) illustrates the situation described in observation (8):

Example (24):

Vysoký činitel … se snaží přimět Palestince k ústupkům na poslední chvíli. (PDT, blc02zu.fs, shortened)

[Top – official – … – tries – (to) force – (the) Palestinians – (to) concessions – at last moment.]

[The senior official … tries to force the Palestinians to retreat at the last moment.]

The noun *Palestince* (N_i) and prepositional group *k ústupkům* (Pg_j) are analysed as Addressee and Pat, i.e. as verbal modifiers. According to the observation (8) the Pg *na (poslední) chvíli* (Pg_k) cannot be treated as an attribute of the first noun *Palestince* (N_i); it is analysed either as an attributive modifier of the preceding Pg *k ústupkům* (Pg_j) or as a (temporal) modifier of the verb *přimět*.

□

IV.3.2. 'Clitic position'

The position of clitics in the surface representation of a sentence is another feature important for Pg-disambiguation. A clitic – usually standing at the 'second position'

(Wackernagel's position) in a sentence – can contribute to disambiguation of the syntactic articulation of a sentence.

Observation (9):

The clitic can separate one noun group (or prepositional group) from the other one, i.e. if there are two noun or prepositional groups in a particular sentence separated by a clitic (N_i/Pg_i – *clitic* – N_j/Pg_j) then the second N_j/Pg_j cannot be treated as an attributive modifier of the first N_i/Pg_i (they are mutually syntactically independent).
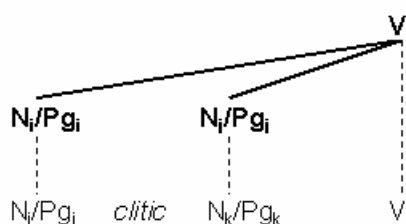

Fig.13: The possible dependency in N_i/Pg_i – *clitic* – N_k/Pg_k chain

□

The following examples (25), (26) and (26') illustrate the contribution of the reflexive pronoun *se* to the Pg-disambiguation:

Example (25):

 *… vláda a následně Parlament ČR **se** po dlouhých odkladech pokusí stanovit pro podnikání seriozní a legislativně jasná pravidla hry.* (PDT, bce17zua.fs #42, shortened)

[… – government – and – subsequently – Parliament – (of) CR – *clitic* – after – long – delays – will try – (to) lay down – for – enterprising – sound – and – legislatively – clear – rules – (of) game.]

 [… the government and subsequently the Parliament of the CR after long delays will try to lay down sound and legislatively clear rules for enterprise.]

There are two rich noun groups in ex. (25), *vláda a Parlament ČR* and *po (dlouhých) odkladech*, separated by the reflexive pronoun *se*. According to observation (9) they are syntactically independent. In fact, both of them modify the verb *pokusí* (*vláda a Parlament ČR* is its Act, *po (dlouhých) odkladech* serves as temporal modifier).

□

Examples (26) and (26'):

(26)   *Pro vytváření generátorů impulsů **se** v číslicových systémech používá oscilátorů.* ([Uhlířová, 1987])

[For – creating – (of) generators – (of) impulses – *clitic* – in – numeric – systems – are used –

oscillators.]

[The oscillators are used in the numeric systems for the creating of generators of impulses.]

In (26) there are two Pg's (*pro vytváření generátorů impulsů* and *v číslicových systémech*) separated by reflexive pronoun *se*, both of them modifying the verb *používá.*

(26') *Pro vytváření generátorů impulsů v číslicových systémech **se** používá oscilátorů.*

[For – creating – (of) generators – (of) impulses –in – numeric – systems – *clitic* – are used – oscillators.]

[The oscillators are used for the creating of generators of impulses in the numeric systems.]

On the other hand, in (26') the first Pg (*pro vytváření generátorů impulsů*) is modified by the second one (*v číslicových systémech*).

□

IV.4. Semantic Features

There are approaches to the syntactic analysis of natural language using only purely syntactic information (for Czech e.g. the Robust Parser mentioned above). The avoidance of semantic information is very well substantiated there by the aims the analysis serves for – the grammar-checking or robustness (i.e. capability to handle also syntactically ill-formed sentences), for example.

On the other hand certain type of semantic information is incorporated in many other automatic parsing systems. For Czech e.g. in RUSLAN, the Czech-to-Russian machine translation project developed in the late eighties, see [Oliva, 1989], or in the grammar checker developed by K. Oliva in LATESLAV project (Language Technologies for Slavic Languages, see [Oliva, 1996]).

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. We want to discuss here a possibility of using such (or slightly modified) semantic features together with semantic features of verbs and prepositions for the purposes of Pg-disambiguation.

If the Pg in a particular sentence is identified neither as a valency modifier nor as a member of 'second order' valency frame of a verb, of some noun or adjective which appear in a 'suitable' position in the sentence (see section IV.1. about word order patterns), then it is time to:

• take into account the word order criteria (see section IV.1.) and formal criteria (see section IV.3. about 'separation principle' and position of clitics) and

• 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 'main' noun in the Pg.

(As was pointed out above, if one of the mentioned valency frames contains some free modifiers, especially temporal and local, the use of the semantic criteria even for identification of these valency modifiers is obvious).

Let us introduce and discuss some examples of semantically based rules for chains of noun and prepositional group with prepositions *na* and *v* (prepositions with possible local or temporal meaning). We will concentrate here on nouns indicating real objects (bearing the feature 'concrete'), and on nouns with features 'time' or 'local', because the relevant rules are the most evident.

Rule (1):

If there is a chain consisting of a noun and a prepositional group $N_i - Pg_j$ (N_i with the semantic feature **'concrete'**, Pg_j consisting of the preposition *na* or *v* and noun N_j in the Locative case with the semantic features **'local'** plus **'proper name'** (one immediately follows the other)) then Pg_j is preferably a non-congruent attributive modifier of N_i.

□

Example (27) illustrates the situation described in rule (1).

Example (27):

Vypsal výběrové řízení na budovu na Brusnici. (PDT, bm102zua.fs #18, shortened, modified)

[(He) advertised – tender – for – building – at – Brusnice.]

[He opened a tender for a building at Brusnice.]

There are two candidates for the noun governing the Pg *na Brusnici* – nouns *(výběrové) řízení* and *budovu*. As the noun *budovu* bearing feature 'concrete' is immediately followed by the Pg *na Brusnici* ('local' plus 'proper name'), according to the rule (1) this Pg is a local attribute of *budovu*. (In fact *na Brusnici* can be also treated as a local modifier of the verb *vypsal* (lower preference), but this case is not covered by this rule.)

□

Rule (2):

If there is a chain in a sentence consisting of a noun and a prepositional group $N_i - Pg_j$ (Pg_j consisting of the preposition *na* or *v* and N_j, both N_i and N_j with the semantic features **'local'** plus **'proper name'**) then their relation is determined by the cognitive content of the sentence. Such extra-linguistic knowledge can be described only with a very detailed classification of semantic features:

• if N_i is 'wider' and N_j is 'closer' then the two nouns are mutually syntactically independent

• if N_i is 'closer' and N_j is 'wider' then the analysis of Pg_j as non-congruent attributive modifier of N_i is preferred.

□

Examples (28) and (28'):

(28) *Nechtěný pobyt na čerstvém vzduchu nabízí svým žákům Svobodná speciální škola J. A. Komenského v Praze na Kampě.*   (PDT, bm102zua.fs #15, second part)

[Involuntary – stay – on – fresh – air – offers – (to) its – pupils – Free / Independent – Special School – (of) J.A. Komenský – in – Prague – on – Kampa.]

[Independent Special School of J.A. Komenský in Prague on Kampa Island offers the involuntary stay on fresh air to its pupils.]

In this case *(v) Praze* and *na Kampě* (in this order; Kampa is a part of Prague) are sister nodes. (The identification of their common noun governor *škola* is implied by the analysis of chain of preceding nouns. The possibility that the Pg *na Kampě* can be treated also as a verbal modifier is not covered by this rule.)

(28') *… Svobodná speciální škola J. A. Komenského na Kampě v Praze.*

[… – Free / Independent – Special School – (of) J.A. Komenský – on – Kampa – in – Prague.]

[Independent Special School of J.A. Komenský on Kampa Island in Prague …]

On the other hand – if *(na) Kampě* precedes the Pg *v Praze* then *v Praze* is preferably treated as an attributive modifier of *(na) Kampě*.

□

The preceding examples (28) and (28') show clearly that such a semantic characterisation of object in the world cannot be embraced by any automatic system and so such type of rules cannot be implemented. This implies that the automatic procedure can cover at most very specific subparts of the world; for the general case the analysis must preserve both above mentioned syntactic structures:

• the first noun N_i is modified by the Pg_j

• N_i and Pg_j are independent on each other.

Discussing here some examples of rules based on semantic features of particular lexical items we must conclude that the possibility of exploitation of such type of semantic information for Pg-disambiguation is disputable and one must be very careful if he/she wants to use it. According to our opinion 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 specifying particular free modifiers (but this is outside our topic). In case of morphemic ambiguity of prepositional groups the rules based on semantic features can be used 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.

V.   ARRANGEMENT OF CRITERIA

The criteria for Pg-detection as they were described in section IV. work on data provided with full morphological analysis. We presuppose that every word of an examined sentence bears complex syntactic and semantic information extracted from lexicon – first of all, every noun, verb and adjective must contain (besides its POS and morphological characteristics) all possible valency frames – the 'first order' as well as the 'second order' ones. Nouns and prepositions are to be marked with their semantic features. In addition to this type of information, possible verbonominal collocations must be identified and equipped with the valencies, frozen collocations must be composed. Before starting the proper analysis of Pg-modifiers we assume the local phenomena – as analytic verb forms and collocations of simple noun groups with numerals – are processed. Some estimation of boundaries of clauses is also presupposed.

We have already stated that our approach to automatic analysis is based on the notion of analysis by reduction. The analysis by reduction is defined as step by step simplification of an examined sentence. Each step consists of deleting and possible rewriting of some words so that correctness of the sentence is preserved[5]. Intuitively, all dependent words must be deleted before their governor is reduced (otherwise the requirement of correctness preserving is not

---

[5] The changes in morphological categories are typical for rewriting, especially in number and gender. They serve as a means for correctness preserving in the course of deleting – e.g. in (22') the agreement between subject and predicate is kept with rewriting.

fulfilled, as is shown in examples (22') and (8))[6]. In each phase any of the mutually independent words can be deleted and this fact causes the existence of several branches of reduction.

Let us show the mechanism of reduction on the following example:

Example (22'): (modification of sentence (22))

Zákon a nařízení dbají na nájemníkovu ochranu. (PDT, bl101js.fs #18, modified)

[Law – and – prescription/prescriptions – look – after – tenant's – protection.]

[The law and prescription/prescriptions look after the tenant's protection.]

The possible reduction steps can be illustrated by the following scheme:
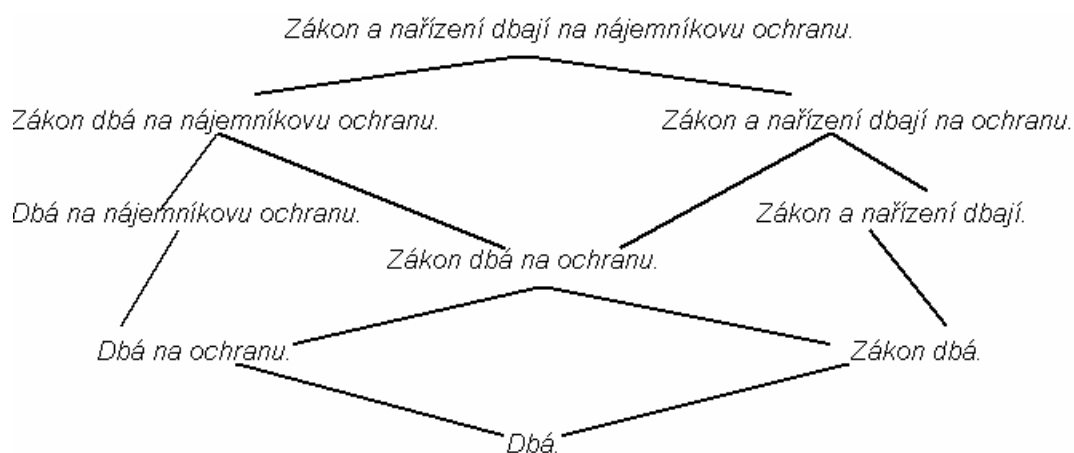

Fig.14: The possible steps during analysis by reduction.

The original sentence can be simplified to *Zákon a nařízení dbají na ochranu.* (word *nájemníkovu* is deleted). The second step can consist of deleting of a group *a nařízení* and rewriting *dbají* (3rd person, plural) by *dbá* [looks after] (3rd person, singular): the sentence *Zákon dbá na ochranu.* is obtained (by deleting *a nařízení* the number of subject nominal group is changed, thus the number of predicate must also be changed). And we can continue in this manner. (The branches of reduction not captured in the scheme are excluded because they do not preserve the condition of correctness).

□

Routine:

During every possible step of analysis in which Pg is reduced all criteria defined in section IV. are applied. If two (or more) possible governors of Pg are detected (either by means of valency requirements or word order patterns) the analysis splits into several branches.

• Word order pattern of simplified sentence is specified.

---

[6] Mel'čuk in [Mel'čuk, 1988] clarifies criteria for the distinction between the governor and the dependent member of dependency relation (from the strictly syntactic point of view) in terms of the passive syntactic

• All <u>valency requirements</u> of autosemantic verbs, nouns and adjectives are examined. With respect to word order limits (as they are formulated in section IV.1. in observations (1)-(4)) every suitable valency slot is filled in by the tested Pg. If some verbonominal collocation is detected then valency requirements are satisfied regardless of word order. Possible new structures are evaluated according to the reliability of valency frames (see section IV.3).

• Information following from <u>word order pattern of simplified sentence</u> is applied. All possible dependencies are marked. New structures are of lower preference than the ones described in the preceding point.

• <u>Semantically based criteria</u> are applied serving for evaluation of existing structures.

• The <u>formal criteria</u> are checked <u>with regard to the original sentence</u>. If any new structure violates these criteria then it is excluded from further analysis.

• If at least one new structure satisfies formal criteria then Pg is deleted. Otherwise this branch of reduction analysis is excluded because it does not preserve the condition of correctness.

□

Let us illustrate on the next examples how the criteria are treated during analysis by reduction:

Example (22): (continuation)

Zákon dbá na ochranu nájemníků. (PDT, bl101js.fs #18, shortened)

[Law – looks – after – protection – (of) tenants.]

[The law looks after the protection of tenants.]

The analysis splits into several branches. The reduction of the Pg *na ochranu* starts in several stages:

1. *Zákon dbá na ochranu nájemníků.* (Nothing has been deleted yet.)

The word order pattern is specified as   N – V – Pg – N. There are two valency requirements – *dbá na ochranu* (the 'first order', *dbát* with frame Act / Pat (*na*+Acc) is picked up[7]) and *zákon na ochranu* (the 'second order') but the second one is excluded because of the word order limits. No verbonominal collocation is detected. No new structure is obtained from the word order pattern. No semantic criterion is applied. No formal criterion is violated with respect to the original sentence. As one valid structure is obtained, Pg *na ochranu* can be deleted (the corresponding Patient slot of verb is marked as being filled in) and reduction

---

valency, or distribution.

[7] There is another valency frame of the verb *dbát* consisting of Actor and Patient (in Genitive case). This frame is picked up by other branch of analysis where the noun *nájemníků* is detected as a Patient.

continues with the next step. (But this branch will fail in following analysis, since the nominal group *(na) ochranu nájemníků* cannot be restored).

2. *Zákon dbá na ochranu.* (The noun *nájemníků* (fills in a valency slot of the noun *ochrana*) has been deleted.)

The word order pattern is specified as   N – V – Pg – N. There are two valency requirements – *dbá na ochranu* (the 'first order', *dbát* with frame Act / Pat (*na*+Acc) is picked up) and *zákon na ochranu* (the 'second order') but the second one is excluded because of the word order limits. No verbonominal collocation is detected. No new structure is obtained from the word order pattern. No semantic criterion is applied. No formal criterion is violated with respect to the original sentence. As one valid structure is obtained, Pg *na ochranu* can be deleted (the corresponding Patient slot of verb is marked as being filled in) and reduction continues with the simplified sentence *Zákon dbá.* (This branch leads to successful analysis.)

3. *Dbá na ochranu nájemníků.* (The noun *zákon* (Act) has been deleted.)

The word order pattern is specified as   V – Pg – N. There is one valency requirement – *dbá na ochranu* (the 'first order', *dbát* with frame Act / Pat (*na*+Acc) is picked up). No word order limit is violated. No verbonominal collocation is detected. No new structure is obtained from the word order pattern. No semantic criterion is applied. No formal criterion is violated with respect to the original sentence. As one valid structure is obtained, Pg *na ochranu* can be deleted (the corresponding Patient slot of verb is marked as being filled in) and reduction continues with the next step. But this branch will fail in following analysis (the nominal group *(na) ochranu nájemníků* cannot be restored).

4. *Dbá na ochranu.* (The nouns *zákon* (Act) and *nájemníků* (fills in a valency slot of the noun *ochrana*) have been deleted.)

The word order pattern is specified as   V – Pg. There is one valency requirement – *dbá na ochranu* (the 'first order', *dbát* with frame Act / Pat (*na*+Acc) is picked up). No word order limit is violated. No verbonominal collocation is detected. No new structure is obtained from the word order pattern. No semantic criterion is applied. No formal criterion is violated with respect to the original sentence. As one valid structure is obtained, Pg *na ochranu* can be deleted (the corresponding Patient slot of verb is marked as being filled in) – the single correct sentence *Dbá.* is obtained, the analysis is successful.

The second and the fourth branch lead to successful analysis, the Pg *na ochranu* is treated as verbal modifier (Patient).

□

Example (8): (continuation)

..., rovná na ramínku vystavený kabát. ([Panevová,1998a])

[… – arranges – on – hanger – exposed – coat.]

[… (she) arranges an exposed coat on a hanger. / … (she) arranges a coat exposed on a hanger.]

The analysis splits into several branches. The reduction of the Pg *na ramínku* starts in several stages:

1. *Rovná na ramínku vystavený kabát.* (Nothing has been deleted yet.)

The word order pattern is specified as   V – Pg – Adj – N. There is no valency requirement. Two structures follow from the word order pattern – the Pg *na ramínku* can modify either the verb *rovná* or the adjective *vystavený*. No semantic criterion is applied. No formal criterion is violated with respect to the original sentence. As two valid structures are obtained, Pg *na ramínku* can be deleted and reduction continues with the simplified sentence *Rovná vystavený kabát.* (This branch leads to two successful analyses.)

2. *Rovná na ramínku kabát.* (The adjective *vystavený* (congruent attribute of the noun *kabát*) has been deleted.)

The word order pattern is specified as   V – Pg – N. There is no valency requirement. One structure follows from the word order pattern – the Pg *na ramínku* modifies the verb *rovná*. No semantic criterion is applied. No formal criterion is violated with respect to the original sentence. As one valid structure is obtained, Pg *na ramínku* can be deleted and reduction continues with the simplified sentence *Rovná kabát.* (This branch leads to a successful analysis.)

3. *Rovná na ramínku vystavený.* (The noun *kabát* (Act or Pat) has been deleted.)

The word order pattern is specified as   V – Pg – Adj. There is no valency requirement. Two structures follow from the word order pattern – the Pg *na ramínku* can modify either the verb *rovná* or the adjective *vystavený*. No semantic criterion is applied. No formal criterion is violated with respect to the original sentence. As two valid structures are obtained, Pg *na ramínku* can be deleted and reduction continues with the simplified sentence *Rovná vystavený.* However, this branch will fail in the subsequent analysis (the nominal group *vystavený kabát* cannot be restored).

4. *Rovná na ramínku.* (The noun *kabát* (Act or Pat) and its congruent attribute *vystavený* have been deleted.)

The word order pattern is specified as   V – Pg. There is no valency requirement. One structure follows from the word order pattern – the Pg *na ramínku* modifies the verb *rovná*. No semantic criterion is applied. No formal criterion is violated with respect to the original

sentence. As one valid structure is obtained, Pg *na ramínku* can be deleted – a single correct sentence *Rovná.* is obtained, analysis is successful.

The first, the second and the fourth branch lead to successful analyses, the Pg *na ramínku* is treated either as free verbal modifier – *rovná na ramínku* or as free adverbial modifier *vystavený na ramínku*. These two readings have the same preference.

□

VI. CONCLUSION

In this article we have illustrated the possibility of searching for rather complicated structures in the Prague Dependency Treebank, namely structures containing prepositional group in such a position in a sentence that the suspicion exists that there is a morphemic ambiguity of Pg in the sentence.

On a relatively rich sample of 'suspicious' sentences obtained from PDT we formulate criteria for Pg-disambiguation. Four types of criteria are investigated: criteria based on word order constraints, on valency requirements of various parts of speech, on semantic features of single words and in addition some formal criteria are stated.

The formal criteria – concerning word order restrictions, the position of the clitics and the separation principle – are of a high reliability. The application of valency information is relatively very safe though some sentences were found where the valency of nouns led to an inadequate structure. On the other hand, the advantage of rules based on semantic features is disputable, they can be used only as supplementary criteria, if at all.

The last section is devoted to the analysis by reduction. The mechanism of this type of analysis is illustrated and the treatment of the proposed criteria is shown.

References

J. Allen: Natural Language Understanding. The Benjamin/Cummings Publishing Company, 1987.

A. Bémová a kol.: Anotace na analytické rovině: Návod pro anotátory. ÚFAL Technical Report TR-1997-03.

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

J. Hajič: 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 1998, pp. 106-132.

J. Hajič, E. Hajičová, J. Panevová, P. Sgall: Specifikace stromových struktur na tektogramatické úrovni značkování a základní kroky překladu z analytické na tektogramatickou úroveň. Technical report of the project GAČR 405-96-0198, Feb. 1999.

E. Hajičová, J. Panevová: Valency (Case) Frames of Verbs. In: Contributions to Functional Syntax, Semantics, and Language Comprehension (ed. P. Sgall), Prague, Academia and Amsterdam, John Benjamins, 1984, pp. 147-188.

T. Holan, V. Kuboň, M. Plátek: Formal tools for separating syntactically correct and incorrect structures. Technical report MFF UK, Charles University, Prague, 1997.

P. Jančar, F. Mráz, M. Plátek, J. Vogel: On Monotonic Automata with a Restart Operation. Journal of Automata, Languages and Combinatorics, No. 4, 1999, to appear.

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

V. Kuboň, T. Holan and M. Plátek: A Grammar-Checker for Czech. ÚFAL Technical Report TR-1997-02.

V. Mathesius: Řeč a sloh. In: Čtení o jazyce a poesii (eds. B. Havránek, J. Mukařovský), Družstevní práce, Praha, 1942, pp. 10-100 (2nd edition).

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

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

K. Oliva: A Parser for Czech Implemented in Systems Q. In: Explizite Beschreibung der Sprache und automatische Textbearbeitung, MFF UK, Praha, 1989.

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

J. Panevová: Nesoglasovannoe opredelenie s točki zrenija analiza dlja mašinnogo perevoda. Prague Studies in Mathematical Linguistics 1, Academia, Prague 1966.

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

J. Panevová: Formy a funkce ve stavbě české věty. Academia, Praha, 1980.

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

J. Panevová: 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, 1994, 223-243.

J. Panevová: Funkční styly a automatické zpracování jazyka. In: Česká slavistika. České přednášky pro XII. mezinárodní sjezd slavistů, Krakov 1998 (ed. H. Bláhová, S. Wollmann a kol.), Slavia, Slovanský ústav AV ČR, 1998, pp. 161-167.

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

J. Panevová, P. Sgall: On Some Issues of Syntactic Analysis of Czech. PBML 34,1980, pp. 21-32.

J. Panevová, M. Straňáková: 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), Lecture Notes in Artificial Intelligence vol.1692, Springer, 1999, pp. 50-55.

P. Piťha: On the Case Frames of Nouns. Prague Studies in Mathematical Linguistics 7, Academia, Prague 1981, pp. 215-224.

P. Piťha: K otázce valence u adjektiv. SaS 43, 1982, pp.113-118

M. Plátek, T. Holan, V. Kuboň, J. Hric: Grammar Development & Pivot Implementation. JRP PECO 2824 Language Technologies for Slavic Languages, Final Research Report, Prague, 1996.

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

P. Sgall a kol.: Úvod do syntaxe a sémantiky. Academia, Praha, 1986.

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

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

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

H. Skoumalová: Derived frames and the lexicon. In: Issues of Valency and Meaning. Studies in Honour of Jarmila Panevová (ed. E. Hajičová), Karolinum, Charles University Press, Prague 1998, pp.154-168.

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

Vl. Šmilauer: Novočeská skladba. SPN, Praha, 1966 (2nd edition).

L. Uhlířová: Knížka o slovosledu. Academia, Praha, 1987