Selected Types of Pg-Ambiguity; Processing Based on Analysis by Reduction

Markéta Straňáková

Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic
marketa.stranakova@ufal.mff.cuni.cz

Abstract. A need for the detection of ambiguous prepositional groups (Pg's) is discussed in the paper. Criteria for automatic Pg-disambiguation are presented, with special accent on word order patterns and verbonominal collocations, the main source of non-projective constructions caused by non-congruent Pg-modifiers of nouns. Usefulness of the proposed criteria and their ordering within an analysis by reduction is introduced.

1 Description of the Problem

Based on the stratificational, dependency-oriented Functional Generative Description of Czech (FGD), four types of ambiguity can be distinguished according to the level of language description containing units which introduce the possibility of two (or more) readings of the sentence.

With prepositional groups (Pg's), the ambiguity defined as an asymmetric relation between the level of surface structure and the adjacent morphemic level is frequent, i.e. the situation when two or more possible functions on the level of the surface representation of the sentence are expressed by the same morphemic form (cf. morphemic ambiguity in the terms defined in [4]). In phrase structure oriented formalisms the corresponding problem is known as PP-attachment, see [1].

From a strictly syntactic point of view a Pg can modify (depend on) a verb, a noun or an adjective (typically deverbal) in the sentence, as in the following example.

Example (1): V období, kdy prudce poklesl zájem na domácím trhu, 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 sharply fell, the factory managed to export part of its production.)
Analyzing the Pg *na domácím trhu* [on (the) domestic market], the valency frame of the noun *zájem* [demand] is important.

zájem ... Act (Gen/Adj$_{pos}$) Pat (*na*+Loc/*o*+Acc)

Thus, the Pg *na domácím trhu* [on (the) domestic market] can be treated either (a) as a free modifier (Locative) belonging to the verb *poklesnout* [to fall], or

(b) as an attribute of the noun *zájem* [demand] - either as (ba) a valency modifier (Patient) or (bb) a free modifier (Locative). (The valency slot of Pat with *zájem* [demand] is occupied by a General Participant in cases (a) and (bb).)

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

The appropriate (human) analysis of a sentence is based on a wide range of syntactic, but also semantic and even pragmatic knowledge. We try to detect the types of input information the automatic procedure can take into account and to determine the criteria both for the detection of Pg-ambiguity and for Pg-disambiguation.

Prague Dependency Treebank (PDT) serves as a basic source of data for Czech.

2 Proposal of the Criteria

We presuppose the existence of a lexicon with detailed information containing complete valency frames for verbs, nouns and also adjectives. The verbonominal collocations must be also kept in the lexicon with their valencies. Optionally, the semantic features of nouns and verbs can be used if they are available. The procedure proposed operates on data provided by full morphological analysis.

2.1 Word Order Patterns (WOP's)

The surface word order serves as the basic purely grammatical clue to the Pg-ambiguity detection. Typically, Pg satisfies word order constraints, which have been embodied in WOP's.

Basic unambiguous WOP's.

- V immediately followed by Pg (such V is the Pg governor);
- Pg at the very beginning of a clause (V is the prototypical Pg governor);
- Adj and Pg are separated by any verb form (V is the Pg governor).

These WOP's are relevant for Pg's in prototypical positions. The analysis of prolific non-projective constructions caused by non-congruent Pg-modifier of noun is discussed in the subsection 2.2.

Basic suspicious WOP's.

- a sequence of N's at the beginning of a clause followed by Pg;
- a sequence of V, N and Pg (in this order);
- Pg between N and V (from the left to the right);
- Pg between V and Adj (V - Pg - Adj - N) or Pg being followed by an adjective, a noun and a verb (in this order) (Pg - Adj - N - 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).

WOP's with a preferable reading.

- N being followed by a congruent Adj and Pg (in this order) (then Pg is preferably treated as belonging to such Adj);
- Pg between N and a congruent Adj (then Pg is preferably treated as belonging to such Adj).

There are some regular exceptions to these rules, as the specific contrastive position of Adj or collocations characteristic for scientific terminology.

2.2 Verbonominal Collocations

A verbonominal collocation can be described as a collocation of a noun and a verb (often with very general meaning) where the noun part has a valency slot filled in by Pg. With verbonominal collocations, the word order constraints may be violated and thus the WOP's defined in subsection 2.1 cannot be taken into account.

Two types of verbonominal collocations can be distinguished; both of them must be kept in the lexicon (among the entries of the noun item).

- A verbonominal collocation consists of a verb with very general meaning and of a noun with valency requirements (type (1), *mít šanci/peníze na co* [to have a chance/money for st.], *podat stížnost na koho/co* [to submit (a) complain of sb./st.]).
- Both a noun and a verb constituting a verbonominal collocation have valency slots requiring the examined Pg (type (2), *klást nároky na koho/co* [to make demands on sb./st.]).

From the purely syntactic point of view there are two systematic possibilities how to analyze such collocations - the Pg can be treated either as an attributive nominal modifier or/and as a verbal modifier, which either is registered in a valency frame, or is its optional free modifier. In agreement with the valency requirements, we propose the following solution.

Solution. Concerning Pg-disambiguation two issues must be solved:

- If only the nominal part of the verbonominal collocation has a valency slot requiring the examined Pg (type (1)) then the Pg is analyzed as an adnominal modifier (in agreement with the valency requirements of the noun). The word order constraints must be relaxed, non-projective constructions may arise (see below).
- For type (2), where both the name and the verb constituting the verbonominal collocation require valency Pg-modifier, WOP's are applied. Two structures are preserved whenever it is possible (the one treating Pg as an adnominal modifier and the other with Pg as a verbal modifier).

Example (2): Finanční prostředky na nákup technického vybavení a učebnic poskytly ... (PDT, 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* [to provide] constitutes a verbonominal collocation together with the noun *prostředek* [means], type (1):

prostředek ... Aim (*na*+Acc, *pro*+Acc, *k*+Gen)

According to the solution proposed the structure with the noun chosen as the Pg governor is preserved (the other one, with the Pg as free verbal modifier (also Aim), is excluded).

In such constructions, the Pg modifies the whole verbonominal collocation and cannot be seen as a purely nominal modifier; this is supported by the changes in syntactic structure of such groups in case of adjectival derivation:

poskytnout (prostředky (na něco)) $\rightarrow$ *prostředky (poskytnuté (na něco))*

[to provide means for st.] [means/sources provided for st.]

The Pg is treated as a nominal modifier in the verbonominal collocation. The other syntactic structure is relevant for the construction with a deverbal adjective: *poskytnutý* (pl *poskytnuté*) [provided] is a congruent attribute following its governing noun *prostředky* [means]; thus, according to the WOP, the Pg *na něco* [for st.] is preferably treated as a modifier of the adjective.

Description of non-projetive constructions. The dependency tree is projective if all nodes (n_i) - lying between a governor node (n_g) and a node (n_d) depending on n_g - depend (directly or indirectly) on this governor node (n_g). Otherwise it is non-projective. (The formal definition of projectivity and non-projectivity can be found e.g. in [3].)

Observations. (valid for verbonominal collocations, type (1))

- A valency modifier of a noun expressed by Pg can precede its noun governor in the surface structure of a sentence. (Analyzing such a Pg modifier WOP's cannot be applied).
- A valency modifier of a noun expressed by Pg and its noun governor can be 'separated' by the verb form, we can get a non-projective structure on the level of surface representation (WOP's cannot be taken into account).
- Free modifiers of a noun expressed by Pg's must follow their governor on the surface level - the WOP's are applied.

Example (3): Na uveřejnění odpovědi má podle něho právo každá fyzická nebo právnická osoba ... (PDT, bmd23zua.fs ♯46, modified))

[on - publication - (of) reaction - has - according to - him - right - every - person - or - legal - entity]

(According to him, every person or legal entity has the right of publication of reaction ...)

The noun *právo* [right] constitutes a verbonominal collocation, type (1) together with the verb *mít* [to have].

právo ... Pat (Gen/*na*+Acc/*k*+Dat/infinitive)

The Pg *na uveřejnění (odpovědi)* [of publication (of reaction)] meets the valency requirement of the noun *právo* [right] resulting in a non-projective construction.

3 Arrangement of the Criteria

An **analysis by reduction** (RA) is a general method that can naturally underlie any automatic parsing procedure.

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

In the next paragraphs we propose the partial use of RA for the detection of Pg-governors. (For the particular description of the criteria used see [6].)

Routine. In every possible stage of analysis in which Pg is reduced all criteria under consideration are applied.

Step 1. The word order pattern of the simplified sentence is specified.

Step 2. All valency requirements of autosemantic verbs, nouns and adjectives are examined. With respect to word order limits (as they are embodied in WOP) every suitable valency slot is filled in by the tested Pg. If some verbonominal collocation, type (1) is detected then the valency requirements are satisfied regardless of the relevant word order limitations. Possible new structures are evaluated according to the reliability of valency frames.

Step 3. The information following from the word order pattern of the simplified sentence is applied. All possible dependencies are marked. New structures are of lower preference than the ones described in the preceding point.

Step 4. Optionally, semantically based criteria can be applied serving for the evaluation of existing structures.

Step 5. 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.

Step 6. If at least one new structure satisfies the formal criteria then Pg is deleted. Otherwise this branch of analysis is excluded because it does not preserve the condition of correctness.

Example (4): Rodiny mají nárok na příspěvek. (PDT, bl103zu.fs, shortened)
[Families have (a) claim to (an) allowance.]
(a) If the verbonominal collocation *mít nárok* [to have (a) claim], type (1) is taken into account the reduction of the Pg *na příspěvek* [to allowance] starts in two stages (see Fig.1, bold arrows). According to the solution proposed above the dependency pair *nárok na příspěvek* [claim to allowance] is created (in agreement with the valency requirement of the noun).
(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 (see Fig.1, dotted arrows). The two analyses satisfy the condition on preserving syntactic correctness, both are successful. However, 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 lies in the lexical entries, therefore it cannot be covered by general rules.

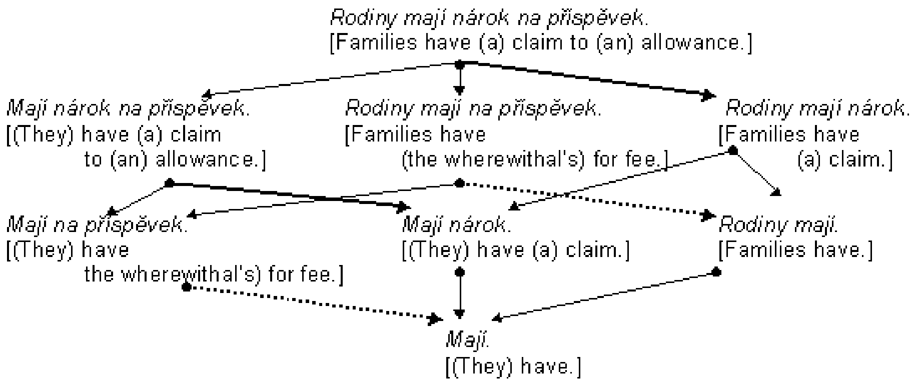

Fig. 1. The possible stages during RA

4 Conclusions

We have tried to illustrate the usefulness of criteria for detection of Pg-ambiguity based on word order patterns and valency requirements of nouns and adjectives. Special interest has been devoted to verbonominal collocations and to their contribution to analysis of non-projective constructions. The last section has been focused on the analysis by reduction. The mechanism of this type of analysis has been illustrated and the treatment of the proposed criteria has been discussed.

Acknowledgement. The research reported on this paper has been supported by the grant VS 96/151.

References

1. Allen, J.: Natural Language Understanding. The Benjamin/Cummings Publishing Company (1987)
2. Holan, T., Kuboň, V., Oliva, K., Plátek, M.: On Complexity of Word Order. Special Issue TAL on Dependency Grammars (to appear)
3. Marcus, S.: Sur la notion de projectivité. In: Zeitschr. f. math. Logic und Grundlagen d. Math. XI (1965) 181-192
4. Panevová, J.: K otázkám homonymie a neutralizace ve stavbě věty. Jazykovedné štúdie **16** (1981) 85-89
5. Sgall, P., Hajičová, E., Panevová, J.: The Meaning of the Sentence in Its Semantic and Pragmatic Aspects. Dordrecht:Reidel and Prague:Academia (1986)
6. Straňáková, M.: Selected Types of Pg-Ambiguity. PBML **72** (1999) 29-58