The Theory of Control Applied to the Prague Dependency Treebank (PDT)

Jarmila Panevová, Veronika Řezníčková, and Zdeňka Urešová
Charles University, Prague, Czech Republic {panevova,rez,uresova}@ufal.mff.cuni.cz

1. Introduction

One of the most difficult issues within corpora annotation on an underlying syntactic level is the restoration of nodes omitted in the surface shape of the sentence, but present on the "underlying" or "deep" syntactic level. In the present paper we concentrate on such type of nodes which are omitted due to the phenomenon usually called grammatical "control" with regard to their respective anaphoric relations. In particular, we extend the notion of control to nominalization and demonstrate how this relation is captured in the Prague Dependency Treebank.

The theory of control is present within Chomsky's framework of Government and Binding (using the terms *verb of control*, *controller* and *controllee*, cf. Chomsky, 1980), but also within many other formal frameworks, e.g. GPSG (Sag and Pollard, 1991) or categorial grammar (Bach, 1979). We analyse this phenomenon within the framework of the dependency grammar, theoretically based on the Functional Generative Description (FGD, cf. Sgall, Hajičová and Panevová, 1986).

In FGD, on the "underlying" or "tectogrammatical" level, control is a relation of an obligatory or an optional referential dependency between a *controller* (antecedent) and a *controllee* (empty subject of the nonfinite complement (= *controlled clause*)). The controller is one of the participants in the valency frame of the governing verb (Actor (ACT), Addressee (ADDR), or Patient (PAT)). The controlled clause functions also as a filler of a dependency slot in the valency frame of the governing verb, being labeled as Patient or Actor. The empty subject of the controlled clause may have the function of different dependency relations to its head word (the infinitive): Actor, or, with passivization of the controlled clause, Addressee or Patient (cf. Koktová, 1992).

2. Capturing of "control" phenomena in the PDT

In the present section we focus on the capturing of the phenomenon of control in the Prague Dependency Treebank (PDT), a three-layer annotated corpus of Czech, basically conceived of in accordance with the theoretical assumptions of the FGD (for more information about PDT cf. Hajič: Tectogrammatical Representation: Towards a Machine Transfer in Machine Translation, this volume).

2.1. Restoration of deletions and capturing of coreferential relations in the PDT

One of the basic principles of annotation of the PDT at the tectogrammatical level concerns also restoration of deletions: in the cases of deletion in the surface sentence, nodes are introduced into the tectogrammatical tree to 'recover' a deleted word. It includes also a restoration of deleted participants of valency frames of verbs. When the nodes deleted in constructions of control are restored, annotators should indicate coreferential relations between the arguments in positions of the controller and the controllee. For labeling these coreferential relations the following attributes (grammatemes) of the general scheme are relevant:

COREF(erence) - the value of this attribute is the lexical value of the antecedent of the given anaphoric node (this node itself may be present on the surface, or deleted)

ANTEC(edent) – the value of this attribute corresponds to the functor of the antecedent with grammatical coreference[1]

CORNUM – refers to the antecedent of the given node[2].

The *Controllee* gets the special lemma *Cor*.

Let us present here some illustrative examples of rather complicated sentences from our annotated corpus that exhibit relations between the arguments in positions of the controller and the controllee.

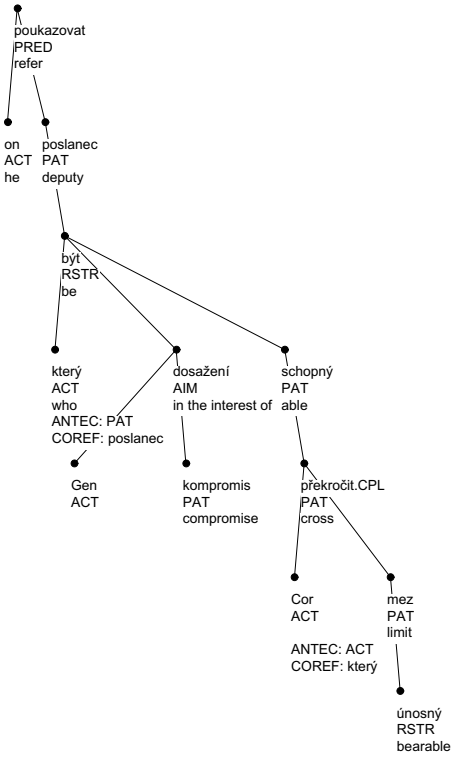
* Supported by the Ministry of Education of the ČR Project LN00A0063.

1. For the difference between the textual and the grammatical coreference see Hajičová, Panevová and Sgall, 2000.

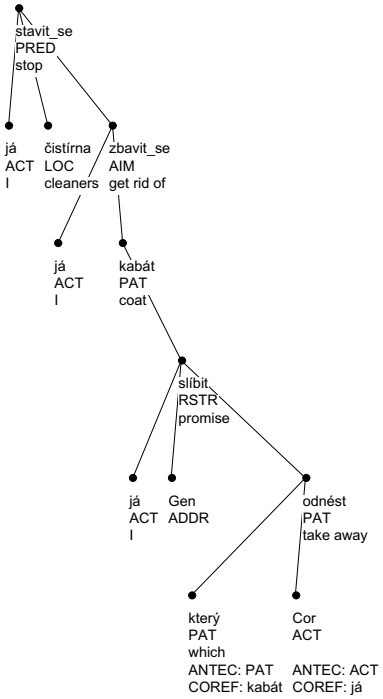
2 Technically, the CORNUM is the only attribute that has to be marked, since the attributes COREF and ANTEC can be then easily extracted from the referred-to node. For the reason of perspicuity we refer to all the three attributes separately.

 Proceedings of the Sixth International Workshop on Tree Adjoining Grammar and Related Frameworks (TAG+6), pp. 175-180. Universitá di Venezia.

(1) Poukazuje na poslance, kteří jsou v zájmu dosažení kompromisu schopni překročit únosnou mez.
'He refers to deputies who are able in the interest of the compromise to cross the bearable limit.'



(2) Musím se stavit v čistírně, abych se zbavil toho kabátu, který jsem slíbil odnést.
'I must stop at the cleaners to get rid of the coat (which) I promised to take away.'



2.2. Survey of views on "control" phenomena with verbs in the FGD

2.2.1. Classification of verbs of control with controlled infinitive clauses

Koktová and Panevová classify the verbs of control according to the type of its valency frame and to the functions of the controlled infinitive clause and the controller in the valency frame of the verb of control (see Koktová, 1992, and Panevová 1986, 1996). According to this classification the following basic groups of verbs of control should be recognized (we leave out here some groups with really rare types of verbs of control, e.g. verbs with the so-called Slavonic Accusative with Infinitive, e.g. *Viděl Karla přicházet* (lit. *He saw Charles to-come*)):

1. The controlled infinitive clause functions as Patient: three groups of verbs of control in Czech can be distinguished, namely verbs in the valency frame of which the Controller is:

i) ACT (e.g. *Jan se bojí zůstat doma sám* (*John is afraid to stay at home alone*))

ii) ADDR (e.g. *Redaktor doporučil autorovi provést několik změn v textu* (*An editor recommended the author to make several changes in the text*))

iii) ACT or ADDR (the verb *slíbit* (*promise*) with the Controller functioning as ACT: e.g. *Jan slíbil matce vrátit se domů před půlnocí* (*John promised his mother to return at home before midnight*); the same verb with the Controller functioning as ADDR e.g. *Rodiče slíbili dětem užít si prázdniny ve stanu u rybníka* (lit.: *The parents promised (their) children to enjoy the holidays in a tent by a lake*))

2. The controlled infinitive clause functions as Actor: especially the "predicate" of control (expressed by a copula with an evaluative or modal adjective) is taken into account (e.g. *Je snadné číst tu knihu* (*It is easy to read the book*))

3. The controlled infinitive clause can have also another function, as cases based on the operation of raising (e.g. *Viktor se zdá být chytrý* (*Viktor seems to be clever*)) and the function of attribute (e.g. *Viktor nesmí propást šanci vyhrát* (*Viktor may not miss the occasion to win*)).

2.2.2. Extension of verbs of control also to the so-called "analytical predicates"

The most typical verbs of control (belonging to the group (1)(i)) are modal verbs (e.g. *moci (can), smět (may), chtít (want), muset (must), mít (have to*)) and so-called "phase verbs" (e.g. *začít (begin), zůstat (stay), přestat (stop*)). While describing the phenomenon of control, it seems to be necessary to extend the understanding of the notion of modal verb also to another synonymous expressions of these verbs. Thus the function of modal verbs is undertaken not only by "modal verbs in the wider sense" (*umět (be able), dovést (know how to do sth), dokázat (manage), zdráhat se (hesitate), odmítat (refuse)* etc.) but also by "analytical predicates" with modal meaning (the verb *mít* (*have*) plus a noun, e.g. *mít schopnost* (lit. *have an ability*)*, dar (*lit. *have a gift / talent), potřebu (have an urge to do sth), příležitost (have an opportunity), šanci (have a chance)*; the verb *být* (*be*) plus a modal adjective, e.g. *být schopen (be able), ochoten (be willing), povinen (be obliged))*.

Also some verbs from other semantic groups of verbs of control can be expressed by some type of "analytical predicate". For example verbs expressing intent, e.g. *hodlat* (*intend), snažit se (try)*, can be paraphrased by predicates *mít v úmyslu (úmysl), záměr* (lit. *have an intention*)*, mít v plánu (plán)* (lit. *have a plan*)*, mít tendenci* (lit. *have a tendency*) etc.*; být připraven (be ready), odhodlán (be determined)* etc. (they belong also to the group (1)(i)). Verbs expressing the meaning *"umožnit někomu udělat něco" (make it possible for somebody to do something)* can be paraphrased by analytical predicates *dát někomu šanci (příležitost) udělat něco* (lit. *give somebody a chance (an opportunity) to do sth*) (these verbs belong to the group (1)(ii)).

2.2.3. Verbs of control with controlled nominalizations

Panevová (1996) deals not only with controlled infinitive verb structures but also with some types of nominalizations where an omission of an argument is also based on the "control" properties of the head (governing) word and must be interpreted as coreferentiality. The group of verbs that offer the possibility for controlled nominalization includes for example verbs from the semantic group of causing a change of a physical and/or mental state, e.g. přisoudit (adjudge), osočit (accuse), podezírat (suspect): Paní podezírá komornou z krádeže stříbrných příborů (The lady suspects the chamber-maid of the theft of silver covers)).

2.3. Nominalizations in constructions of control

The restoration of deletions in PDT includes not only the restoration of all obligatory participants and obligatory free modifications of verbs deleted at the surface shape of the sentence, but also the restoration of obligatory members of valency frames of postverbal nouns and adjectives formed by the process of nominalization.

2.3.1. From verbs to nouns

By nominalizations we understand:

a) Nouns derived from verbs by productive means (e.g. *rozhodnutí (decision making)*, *obžalování* (*accusing*) or nouns derived from verbs by non-productive means or by the zero suffix (e.g. *rada* (*advise*), *slib* (*promise*))

b) Nouns derived from the predicative adjective (e.g. *on je schopen udělat* (*he is able to do sth*) → *jeho schopnost napsat knihu (his ability to write a book)*, *on je povinen udělat* (*he is obliged / required to do sth*) → *jeho povinnost vydat majetek (his duty / obligation to release possession)*

c) Deverbative adjectives, it seems that only predicative deverbative adjectives can occur with control (e.g. *dívka je schopna studovat* (*the girl is able to study*)→ *dívka schopná studovat (a girl able to study) , osoba je povinna platit daně (the person is obliged to pay taxes) → osoba povinná platit daně (a person obliged to pay taxes)*

d) Nouns which were a part of an analytical predicate (e.g. *Petr má šanci vyhrát (Peter has a chance to win*) → *Petrova šance vyhrát* (*Peter's chance to win*)*, Petr má právo odvolat se (Peter has a right to appeal)* → *Petrovo právo odvolat se (Peter's right to appeal)*.

Some of the nouns derived from this type of analytical predicates, especially from those with the meaning of intent, do not express grammatical coreference, e.g. *nápad vydat knihu (an idea to publish a book)* (cf. also Panevová, 1996).

2.3.2. Types of nominalized constructions of control

Considering the possibility of a nominalization of both the governing as well as the dependent verb, we deal with four types of constructions of control:

1. The infinitive clause depends on a finite verb (e.g. *radil nechodit (he advised not to go)*, *slíbil napsat (he promised to write);*

2. The infinitive clause depends on a nominalization of a finite verb (e.g. *rada nechodit* (*an advice not to come*), *slib napsat* (*a promise to write*));

3. The nominalization of the embedded verb depends on a finite verb (e.g. *obvinil někoho z vyvolání problému (he charged a person with a raising of a problem), vyžadoval odpuštění daní (he claimed exemption of the taxes))*;

4. The nominalization of the embedded verb depends on a nominalization of a finite verb (e.g. *obvinění z vyvolání problému (an accusation of a raising of a problem), snaha o podplacení (an attempt for corruption)).*

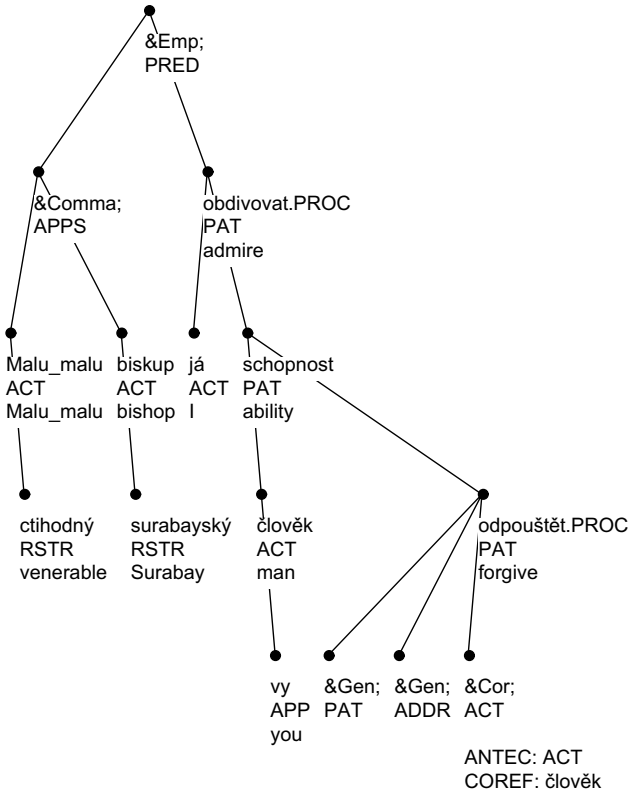
However, it is necessary to say that not all groups of verbs of control mentioned in section 2.2.1. allow for its nominalization or for a nominalization of its controlled infinitive clause:

- Verbs of control from the groups (1)(i), (ii) and (iii) may occur in all four types of constructions of control (e.g. verbs *slíbit* (*promise*), *vyžadovat* (*require, claim*), *snažit se* (*try*): *slíbit napsat* (*to promise to write*), *slib napsat* (*a promise to write*), *slíbit napsání* (*to promise writing*), *slib napsání* (*a promise of writing*)
- Verbs of control from the group (2) allow only for the nominalization of the dependent verb (*Je snadné číst tu knihu* (*It is easy to read the book*) - *Četba této knihy je snadná* (*The reading of this book is easy*)
- Verbs from the group (3) do not allow nominalization in constructions of control.

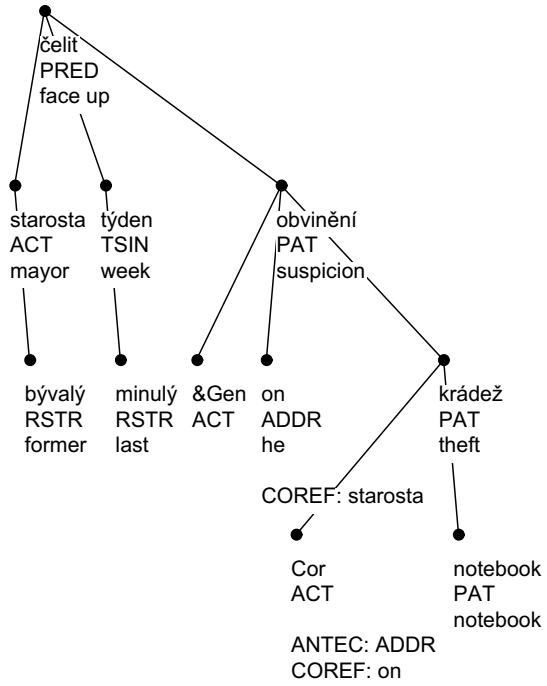
Verbs mentioned in section 2.2.3. may occur only in construction types (3) and (4) (e.g. verbs *podezírat* (*suspect*), *obvinit* (*accuse*): *podezírat z krádeže* (*to suspect of theft*), *podezření z krádeže* (*a suspicion of theft*), but **podezírat krást* (*to suspect to steal*), **podezření krást* (*a suspicion to steal*)).

Let us present here some illustrative examples of nominalized constructions of control from our annotated corpus:

(3) Ctihodný Malu-malu, biskup Surabayský: Obdivuju schopnost Vašich lidí odpouštět.
'The venerable Malu-malu, the bishop of Surabaya: I admire the ability of your people to forgive.'



(4) Bývalý starosta od minulého týdne čelí obvinění z krádeže notebooku.
'The former mayor has been facing up to suspicion of theft of the notebook since the last week.'



2.3.3. Coreferential relations in nominalized constructions

Nominalized constructions retain the same coreferential relations between the Controller and the Controllee which were realized in constructions with the corresponding verbs of control. Thus, e.g. the nominalized constructions of verbs from the group (1)(iii) mentioned in section 2.2.1. offer the possibility for the Controller to be an Actor or an Addressee. These features are illustrated in the following examples:

1. Constructions in which the Actor of the governing postverbal noun and the Actor of the dependent noun (derived from the predicate expressed by a copula with an adjective) are identical:

(5) jeho slib poslušnosti
derived from the construction *slíbil, že bude poslušný* (*he promised to be obedient*)
'his promise of obedience'

The controllee in the valency frame of the dependent noun (i.e. *poslušnost* (*obedience*)) gets the lemma *Cor* and the functor ACT. Its attributes for coreferential relations are filled in by the following values: COREF: *on* (*he*), ANTEC: ACT.

2. Constructions in which the Actor of the dependent noun (derived from the predicate expressed by a copula with an adjective) is identical to the ADDR of the governing postverbal noun:

(6) slib beztrestnosti
derived from the construction *slíbili mu, že bude beztrestný* (*they promised him to be exempt from punishment*)
'a promise of impunity'

The Controllee in the valency frame of the dependent noun (i.e. *beztrestnost* (*impunity*)) gets the lemma *Cor* and the functor ACT. Its attributes for coreferential relations are filled in by the following values: COREF: *on* (*he*), ANTEC: ADDR.

3. Conclusion

In the present paper we sum up how the "control" phenomenon is treated in the framework of the FGD and demonstrate how annotators capture the control properties in the PDT. We also presented the extension of the notion of verbs of control to the so-called "analytical predicates", but especially to the nominalized constructions of control. We showed that the nominalized constructions retain the same coreferential relations between the Controller and the Controllee as those realized in constructions with the respective verbs of control.

References

Bach, Emmon. 1979. Control in Montague grammar. Linguistic Inquiry 10: 515-531.

Chomsky, Noam. 1980. On Binding. Linguistic Inquiry 11: 1-46.

Hajič, Jan. 2002. Tectogrammatical Representation: Towards a Machine Transfer in Machine Translation. In: Proceedings of TAG+6, 2002, Venice, Italy, this volume.

Hajičová, Eva, Jarmila Panevová, and Petr Sgall. 2000. Coreference in Annotating a Large Corpus. In: Proceedings of LREC 2000, 1: 497-500, Athens, Greece.

Koktová, Eva. 1992. On New Constraints on Anaphora and Control. Theoretical Linguistics 18: 102-178.

Panevová, Jarmila. 1986. The Czech Infinitive in the Function of Objective and the Rules of Coreference. In: *Language and Discourse: Test and Protest* (Festschrift for P. Sgall, ed. by J. Mey), Amsterdam – Philadelphia: John Benjamins, 123-142.

Panevová, Jarmila. 1996. More Remarks on Control. In: *Prague Linguistic Circle Papers*, 2 (eds. E. Hajičová, O. Leška, P. Sgall, Z. Skoumalová). J. Benjamins Publ. House: Amsterdam - Philadelphia, 101-120.

Sag, Ivan, and Carl Pollard. 1991. An Integrated Theory of Complement Control. Language 67: 63-113.

Sgall, Petr, Eva Hajičová and Jarmila Panevová. 1986. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Dordrecht: Reidel and Prague: Academia.