

Reconstructions of Deletions in a Dependency-based Description of Czech: Selected Issues

Eva Hajičová and Marie Mikulová and Jarmila Panevová

Charles University in Prague

Faculty of Mathematics and Physics

Institute of Formal and Applied Linguistics

Czech Republic

{hajicova,mikulova,panevoval}@ufal.mff.cuni.cz

Abstract

The goal of the present contribution is to put under scrutiny the language phenomenon commonly called ellipsis or deletion, especially from the point of view of its representation in the underlying syntactic level of a dependency based syntactic description. We first give a brief account of the treatment of ellipsis in some present day dependency-based accounts of this phenomenon (Sect. 1). The core of the paper is the treatment of ellipsis within the framework of the dependency-based formal multi-level description of language called Functional Generative Description: after an attempt at a typology of ellipsis (Sect. 2) we describe in detail some selected types of grammatical ellipsis in Czech (Sect. 3). In Sect. 4 we briefly summarize the results of our analysis.

1 Treatment of ellipsis in dependency based descriptions of language

There are not many treatments of ellipsis in the framework of dependency grammar. Hudson's original conviction presented in his 'word grammar' (WG, (Hudson, 1984)) was that syntactic theory could stick firmly to the surface with dependency relations linking thoroughly concrete words. Under this assumption, such elements as those for which transformational grammar has postulated deletions, traces or unpronounced pronouns such as PRO and *pro* were part of semantics and did not appear in syntax. In his more recent work, (Hudson, 2007), pp. 267-281 revised this rather extreme position; he presents an analysis of examples of structures such as *You keep talking* (sharing of subjects), or *What do you think the others will bring* (extraction) or case agreement in predicatives (in languages such as Icelandic and Ancient Greek, where adjectives and

nouns have overt case inflection and predicative adjectives agree with the subject of their clause) demonstrating that their description cannot be relegated to semantics. He concludes that covert words have the same syntactic and semantic characteristics expected from overt words and, consequently, he refers to them as to the 'unrealized' words. He proposes to use the same mechanism used in the WG theory: namely the 'realization' relation linking a word to a form, and the 'quantity' relation which shows how many instances of it are expected among the observed tokens. If the quantity of the word is zero then a word may be unrealized. Every word has the potential for being unrealized if the grammar requires this. An unrealized word is a dependent of a word which allows it to be unrealized, thus the parent word controls realization in the same way that it controls any property of the dependent.

One of the crucial issues for a formal description of ellipsis is the specification of the extent and character of the part of the sentence that is being deleted and has to be restored. Already in the papers on deletion based on the transformational type of description it has been pointed out that the deleted element need not be a constituent in the classical understanding of the notion of constituent. A natural question offers itself whether a dependency type of description provides a more adequate specification in terms of a dependency subtree. (Osborne et al., 2012) proposed a novel unit called *catena* defined as a word or a combination of words that is continuous with respect to dominance. Any dependency tree or subtree (complete or partial) of a dependency tree qualifies as a *catena*. The authors conclude that based on the flexibility and utility of this concept, *catena* may be considered as the fundamental unit of syntax and they attempt to document this view by their analysis of different kinds of ellipsis (gapping, stripping, VP ellipsis, pseudogapping, sluic-

ing and comparative deletion, see (Osborne and Liang, 2015)).

The issue of ellipsis as a mismatch between syntax and semantics is most explicitly reflected in those dependency frameworks that work with several levels of syntactic representation. This is the case of the Meaning-Text Theory (MTT) of I. Mel'čuk and the Functional Generative Description (FGD) of P. Sgall.

In the framework of the multilevel approach of MTT the rules for surface syntactic ellipsis are part of surface syntax component and they are defined as "various kinds of reductions and omissions, possible or obligatory in a given context ..." (Mel'čuk, 1988), p. 83). For the surface syntax representation the author distinguishes between zero signs and ellipsis. Zero lexes and lexemes are covered by the term syntactic zeroes (op. c., p. 312) and due to their sign character they are reflected in the dictionary entries. On the other hand, an ellipsis is a rule, i.e. a part of the grammar, "that eliminates certain signs in certain surface contexts." (op. c., p. 326).

2 Treatment of ellipsis in the Functional Generative Description

In the dependency-based theory of the Functional Generative Description (FGD) we subscribe to (see esp. (Sgall et al., 1986)) the treatment of ellipsis is determined by the fact that this theoretical framework works with two syntactic levels of the sentence, namely with a level representing the surface shape of the sentence and the level representing the underlying, deep syntactic structure of the sentence (so-called tectogrammatical level).¹ Simplified examples of representations on these two levels for sentence (1) are presented in Fig. 1.

- (1) Jan se rozhodl opustit Prahu.
John Refl. decided to_leave Prague

In the surface structure representation each element of the sentence is represented by a node of its own (more exactly, by the form given in the dictionary) and no words are added. The dependency re-

¹FGD served as a theoretical background of the annotation scheme of the Prague Dependency Treebank (PDT in the sequel; see (Bejček et al., 2013)). PDT also distinguishes an analytic syntactic level (surface) and a tectogrammatical, deep level. In the present contribution, we discuss deletions from the point of view of the theoretical approach and quote PDT only when necessary for the understanding of the point under discussion. For the treatment of deletions in the PDT see (Hajič et al., 2015).

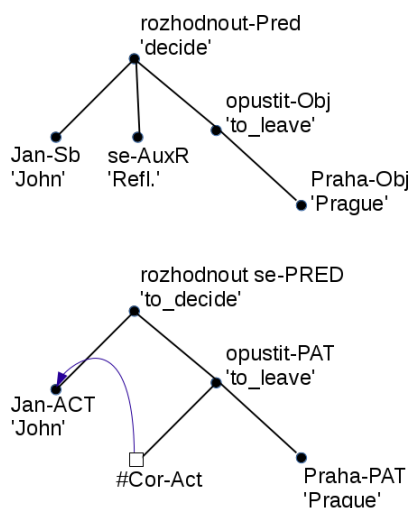


Figure 1: Simplified representations of the sentence (1) *Jan se rozhodl opustit Prahu* [John decided to leave Prague.] on the surface (above) and on the tectogrammatical (below) levels. The arrow indicates the coreferential relation.

lations have the values such as SUBJ, OBJ, ADV etc. In the tectogrammatical tree (TR in the sequel), only autosemantic lexical units are represented by a separate node of the tree; the information carried by the function words in the surface structure is represented in the tectogrammatical structure by means of complex symbols attached to the given node (e.g. the so-called grammatemes of modality, tense, etc. or the subfunctors for the meanings carried by the prepositions etc.). The semantic relation between the head and its modifier(s) is reflected by the functor(s), such as ACT, PAT, ADDR, LOC, CPR, RSTR etc., which are, if needed, supplied by more subtle syntactico-semantic distinctions reflected by the subfunctors.

The issue of ellipsis² concerns the relations between these two dependency trees. It is obvious that for an adequate representation of meaning elements of different dimensions absent on the surface need to be included in the TR. We call these elements ellipsis.

The phenomenon of ellipsis is caused by several factors:

- (i) by the structure of the text (discourse),
- (ii) by grammatical rules or conditions,
- (iii) by an obligatory grammatically determined

²In the present discussion, we use the terms "deletion" and "ellipsis" as synonyms though we are aware that in some frameworks their meanings do not overlap.

surface deletability of an element the presence of which is required by the grammatical system.

Type (i) is called a textual ellipsis, as it is basically connected with the structure of discourse,³ and the types (ii) and (iii) are called systemic (or grammatical) ellipsis; the type (iii) is referred to here as pseudodeletion. In the case of grammatical ellipsis the surface sentences (the "remnants") without the elliptical elements satisfy the conditions for grammatically well-formed structures; however, in order to achieve a representation of the meaning of the sentence these elements have to be filled (often using artificial nodes) in the tree even if the result of the restoration of the deletion may be stylistically awkward or even grammatically hardly acceptable in the surface shape of the sentence. On the borderline between the types (i) and (ii) there is the surface deletion of subject in Czech as a language with the property of a pro-drop language.⁴

3 The FGD treatment of selected types of systemic ellipsis in Czech

As already mentioned above, one of the crucial issues for a formal description of ellipsis is the specification of the extent of the part of the sentence that has to be restored. The extent of the restorations varies from type to type, from the more easily identifiable with the restoration of ellipsis in pro-drop cases to the least identifiable structures to be inserted in cases of deletions in coordination. In our discussion below we will concentrate on four types of systemic ellipsis in Czech with which we intend to illustrate the different possibilities and difficult points of reconstructions; we leave aside deletions in coordinated structures, which is a problem of its own and the discussion of which would go beyond the limits of this contribution.

While in 3.2 – 3.4 the problem how the items absent on the surface are to be reconstructed in TRs (as to their structure and extent), in 3.1 the reconstruction on TR is quite simple, it concerns a single node and it is manifested by the morpho-

³So-called "textual ellipsis" typical for the spoken language and dialogues is left aside here, outside a broader context these sentences may be ungrammatical (as is the second sentence in *Have you finished your manuscript? Not yet completely.*). Their analysis is a subject of studies on discourse structure.

⁴For a detailed classification of ellipsis in Czech, see (Mikulová, 2011).

logical categories of verb. We face here an opposite problem: how to explain the conditions where "pro-dropped" subjects are overtly expressed. In 3.1 we give only several examples with overt subjects in 1st and 2nd person without their deep analysis. By this preliminary picture of the problem we wanted to demonstrate that Czech really belongs to the "pro-drop" class of languages (see Table 1).

3.1 The pro-drop parameter in Czech

Czech belongs to languages of the pro-drop type (called sometimes zero subject or null-subject). Surprisingly, the absence of an overt subject in 1st and 2nd person was not described properly in traditional Czech grammatical handbooks (cf. (Havránek and Jedlička, 1960), p. 300 and in (Karlík et al., 1995), pp. 411–412.). The analysis of this phenomenon is given in more details in contrastive studies, esp. in those comparing Czech and Russian, because these two closely related languages differ as to their pro-drop properties.⁵ Since the examples with missing pronouns of 1st and 2nd person are considered as unmarked for Czech,⁶ while the overt presence of the pronouns in 1st and 2nd person as marked counterexamples, the conditions or requirements for their presence need to be listed. For the 1st person sg the following issues are mentioned in the books quoted above:

(i) the verb forms do not indicate fully the source for the agreement categories (see (2)), (ii) the contrasting position of the pronoun with regard to the other element (see (3)), (iii) the stressed position of the pronoun (often at the beginning of sentence, see (4)), (iv) the pronoun participates in a coordination chain (see (5)), and finally (v) the stylistic feature expressing pleasant or unpleasant emotions (see (6)).⁷

(2) *Já* byl vždycky tak trochu pobuda.

'I have always been a kind of a loungeur.'

⁵A detailed analysis is given in (Isačenko, 1960), Vol 2, pp. 411f.; the author's approach seems to be too radical as to the difference between non pro-drop Russian contrary to the pro-drop Slovak; he proposed to analyse Russian constructions as *Ja splju* [*I am sleeping*] with obligatory subject pronoun *ja* [*I*] as an analytical verb form.

⁶In this section we do not pay an attention to the 3rd person; its position on the scale of deleted elements is different due to its role of anaphora.

⁷The occurrence of pronouns in marked positions in (1) through (11) is denoted by italics; these examples are taken over from the different parts of the Czech National Corpus, namely SYN2010 and SYN2013PUB.

- (3) Byli bohatí, *já* jsem byl chudý.
'[They] were rich, *I* was poor.'
- (4) Ten článek jsem psal *já*.
'The article *I* wrote.'
- (5) Můj přítel a *já* jsme odešli z policejního úřadu.
'My friend and *I* left the police station.'
- (6) *Já* jsem ti, Radku, tak šťastný, že *I* am you, Radek, so happy that už s tebou nemusím hrát. no-longer with you need-not play.
'I am so happy, Radek, that I do not need to play with you any longer.'

The ellipsis of 1st person pl and 2nd sg and pl are not analyzed in the quoted books at all, we present here only several examples of the marked positions untypical for a pro-drop language:

- (7) *My* si na něho počkáme,
We Refl. for him wait,
neuteče nám.
he will not escape us.
- (8) Posekám ti zahrádku a *ty* mi za [I] will cut you garden and *you* me for to vyvenčíš psa.
that will take out dog.
- (9) Vyrozuměli jsme, že právě *vy* jste se s ním stýkala nejčastěji ze všech.
'We have understood that exactly *you* have been meeting him most frequently from all of us.'
- (10) *Ty* nevíš, kdo *já* jsem?
'*You* do not know who *I* am?'
- (11) ...někdo plakal nad čerstvým hrobem a *my* šli a položili ho do hlíny.
'... somebody wept on his fresh tomb and *we* went and put him into the soil.'

In Table 1 we compare the number of sentences with an overt pronominal subject and the number of all sentences with the verb in the form corresponding to this person.⁸ The degree of pro-

⁸The number of occurrences cannot be accurate: the forms *já*, *ty*, *my*, *vy* in nominative could occur in non-subject positions in phrases introduced by *jako* [as]. Both meanings of the pronoun *vy* [you], i.e. the honorific form and the simple plural form would be difficult to distinguish in the corpus without syntactic annotation. However these occurrences are marginal, so that they do not influence the statistics substantially.

corpus	SYN2005	SYN2010	SYN2013 PUB
corpus size (# of tokens)	100M	100M	935M
Verbs in 1 st person sg	1 142 609	1 787 638	8 906 455
Pronoun <i>já</i> [I] is present	77 629	74 922	244 667
non-dropped	6,8%	4,2%	2,7%
Verbs in 2 nd person sg	293 068	496 304	2 966 819
Pronoun <i>ty</i> [you] is present	10 265	17 328	9 779
non-dropped	3,5%	3,5%	0,3%
Verbs in 1 st person pl	635 962	821 381	8 501 392
Pronoun <i>my</i> [we] is present	18 213	19 986	153 275
non-dropped	2,9%	2,4%	1,8%
Verbs in 2 nd person pl	379 487	498 943	1 093 271
Pronoun <i>vy</i> [you] is present	16 596	17 344	65 707
non-dropped	4,4%	3,5%	6,0%

Table 1: Non pro-drop vs. pro-drop sentences

dropness is demonstrated in the 'non-dropped' rows: e.g. in the corpus SYN2005 there are 6,8% sentences within the set of all predicates in 1st person sg where the subject *já* [I] is present (non-dropped).

3.2 Coreference with raising and control verbs as "pseudo-deletions"

With regard to our aim to introduce into the deep (tectogrammatical) representation all semantically relevant information even though not expressed in the surface shape of the sentence, the coreferential units important for the interpretation of the meaning of the sentence in infinitive constructions have to be inserted. Neither speaker nor recipient are aware of any deletion in (12) and (13) (and other examples in this Section), both sentences are fully grammatical.

Thus, for the interpretation of the meaning of (12) it is necessary to know that in (12) Actor (*John*) is identical with absent subject of the infinitive clause, see Figure 1 above, while in (13) the Addressee (*girl-friend*) occupies such an empty position. These elements (indicated in PDT by the lemma #Cor) are needed for the completion of the tectogrammatical structure.

Infinitive clauses with some verbs of control are in particular contexts synonymous with the corresponding embedded clauses (12b), (13b):

- (12) a. Jan se rozhodl opustit Prahu.
'John decided to leave Prague.'
- b. Jan se rozhodl, že (on) opustí Prahu.
'John decided that (he) would leave Prague.'
- (13) a. Jan doporučil přítelkyni přestěhovat se.
'John recommended to his girl-friend to move.'
- b. Jan doporučil přítelkyni, aby se (ona) přestěhovala.
'John recommended to his girl-friend that (she) moved.'

Another argument for the treatment of these structures as deletions is the fact that with some verbs the surface shape of the sentence is ambiguous: thus with the Czech verb *slibovat* [to promise] there are two possibilities of control (the subject of the infinitive may corefer either with the Actor or with the Addressee of the main clause) that have to be captured by the TR. Thus the sentence (14) can be understood either as (15a) with the Actor as the controller or as (15b) with the Addressee as the controller:

- (14) Jirka slíbil dětem jít do divadla.
'George promised the children to go to the theatre.'
- (15) a. Jirka slíbil dětem, že (on) půjde do divadla.
'George promised the children that (he) will go to the theatre.'
- b. Jirka slíbil dětem, že (ony) půjdou do divadla.
'George promised the children that (they) will go to the theatre.'

The specificity of this type of deletion is caused by the fact that the deleted unit – subject (Sb) of the infinitive – cannot be expressed on the surface.

Raising and control constructions belong to the prominent topics of the studies in generative grammar, though different terminology and different solutions are used ((Růžička, 1999), (Przepiórkowski and Rosen, 2005), (Rosen, 2006), (Landau, 2013), to name just a few contributions from the last 20 years).⁹ (Panevová,

⁹ (Růžička, 1999), p.4: "...an infinitival S-complement

1996) and (Panevová et al., 2014) base the solution on the classification of verbs of control according to their controller (examples (12) and (13) represent group 1 and 2 with *Actor (controller) – Sb (controlee)* and *Addressee (controller) – Sb (controlee)*, respectively). The other groups are represented by the Czech verbs *slibovat* [to promise] with two possibilities of control (*Actor - Sb* or *Addressee - Sb*, see (15a), (15b)) and *poslat* [to send] with the control *Patient - Sb* (see (16)).

- (16) Šéf poslal asistenta roznést letáky.
'The boss sent the assistant to distribute the leaflets.'

Our discussion indicates that we have resigned on the difference between raising and control,¹⁰ because according to the analysis of Czech data, the tests (such as passivization, identity or difference in theta-roles, the number of arguments of the head verb) prominently used in generative grammar for English do not function for our data in the same way.

In this Section we wanted to document that phenomena analyzed here and called "pseudo-deletions" are justified to be considered as a type of deletion, as the meaning of infinitive constructions can be explained only by an establishment of explicit pointers of the coreferential expressions between the argument of the governing verb and unexpressed subject of the dependent predicate.

3.3 Special types of "small clauses"

A sequence of two prepositions following one another is excluded in Czech but there are expressions in Czech¹¹ classified in traditional descriptions and dictionaries mostly as prepositions that can be followed by a prepositional noun group.

- (17) Kromě do katedrály půjdou turisté do musea.¹²

creates the problem of reconstituting its empty subject"; (Landau, 2013), p. 9: "... the interpretation of the sentence [*with control*] indicates that there is an additional, invisible argument in the embedded clause, which is coreferential with (found/controlled by) the overt DP."

¹⁰(Landau, 2013), p. 257 concludes his exhaustive analysis of the phenomena analyzed usually under the roof of raising/control by the claim that control "is neither a unitary phenomenon nor a constitutive element of grammatical theory", but rather "a heuristic label only serving to draw our attention to a certain class of linguistic facts".

¹¹Equivalent expressions in other languages (e.g. in Russian), of course, exist, but as far as we know, they do not share the properties we describe for Czech in this Section.

¹²The variant *kromě* + Genitive (*kromě katedrály půjdou*

'Besides to the cathedral the tourists will go to the museum.'

- (18) a. Místo do Uppsaly přijel Jan do Trondheimu.

'Instead at Uppsala John arrived at Trondheim.'

- b. Místo, aby (Jan) přijel do Uppsaly, přijel Jan do Uppsaly, přijel Jan do Trondheimu.
Trondheim.

In our proposal the double functions concentrated in "small clauses" introduced by *kromě*, *místo* [besides, instead of] are differentiated by means of the addition of the missing predicate with the lexical label repeating the lexical value of the governing predicate. The adverbials *do katedrály* (in (17)), *do Uppsaly* (in (18)) depend on the restored node with their proper function of Direction. The expanded representation for (18a) is paraphrased in (18b).

We deal here with examples (17) and (18) in detail, because they document clearly that the (lexically parallel) predicate is missing on the surface. However, there are examples where the preposition *místo* [instead of] is used with its "regular" case rection (Genitive), being sometimes synonymous with the small clause with double prepositions, e. g. (19), (20):

- (19) Místo zavřeného musea(Genitive) navštíví turisté katedrálu.

'Instead of closed museum(Genitive) the tourists will attend a cathedral.'

- (20) Místo manžela(Genitive) doprovodí matku na ples syn.

'Instead of her husband(Genitive) her son will accompany mother to the ball.'

There are two possible approaches how to represent (19) and (20) on TR: In the former case, the

... [besides the cathedral they will go ...] where the expression *kromě* can function as a proper preposition governing genitive case exists in Czech, too, but it is not applicable in all contexts. E.g. *Kromě s přítelem půjde Marie do divadla se sestrou* [lit. Besides with the boy-friend Mary will go to the theatre with her sister] cannot be changed into **Kromě přítele půjde Marie do divadla se sestrou*. [*Besides the boy-friend Mary will go to the theater with her sister.]

expressions *místo muzea/místo manžela* [instead of museum/instead of husband] could be represented as adjuncts of SUBST(itution) directly dependent on the predicate (*visit* or *accompany*, respectively). In the latter case, in order to achieve a symmetric representation of (18) on the one side and (19), (20) on the other, the restored version (with a repeated predicate) will be used. We preferred the latter solution which helps to eliminate an ambiguity such as in (21) paraphrased in (22a) and (22b):

- (21) Místo profesorky kritizoval studenta děkan.

'Instead of the (lady)professor-Gen-F the dean criticized the student.'

- (22) a. Místo aby kritizoval
Instead of that he-criticized

profesorku , kritizoval
the (lady)professor-Acc-F , criticized
děkan studenta.
the dean the student-Acc-F

'Instead of criticizing the lady-professor, the Dean criticized the student.'

- b. Místo aby studenta
Instead of that the student-Acc-F

kritizovala profesorka ,
criticized (lady)professor-Nom-F ,
kritizoval ho děkan.
criticized him the dean.

'Instead of the student having been criticized by the lady-professor, he was criticized by the Dean.'

In the primary meanings of these two sentences in their restored (expanded) versions the noun *profesorka* [lady-professor] after the preposition *místo* [instead of] has the function of the subject (Actor) in (22b), while in (22a) *profesorka* [lady-professor] has the function of object (Patient).

There are additional problems connected with the expression *kromě*. This Czech expression has two meanings corresponding approximately to *besides* (inclusion) and *with exception* (exclusion). At the same time, both have the same syntactic properties. Sentences (23a) and (24a) and their proposed expansions (23b) and (24b) illustrate the two different meanings of structures with *kromě*.

- (23) a. (Tento přímořský hotel nabízí vynikající služby.) Kromě v moři tam můžete plavat (i) v bazénu.
'(This seaside hotel offers excellent services.) Besides in the sea you can swim there (also) in the pool.'
- b. Kromě toho, že tam můžete plavat v moři, můžete tam plavat swim in sea, you-can there swim (i) v bazénu. (also) in pool.

For (24a) we propose the extended tectogrammatical representation as paraphrased in (24b):

- (24) a. Kromě v pondělí můžete navštívit muzeum denně od 10 do 18 hodin.
'With the exception on Mondays you can visit the museum daily from 10 AM till 6 PM.'
- b. Kromě toho, že nemůžete navštívit muzeum v pondělí, můžete navštívit muzeum denně od 10 do 18 hodin.
'With exception of the fact that you cannot visit the museum on Monday, you can visit the museum daily from 10 AM to 6 PM.'

The restored versions of the small clauses serve also as the means how to remove the ambiguities in *kromě*-phrases.¹³ If in the extended version with the restored predicate both predicates are positive or both are negated, the *kromě*-phrases mean inclusion (called Addition in (Panevová et al., 2014)); if one of them is positive and the other negated, the phrases express an exclusion (called Exception in (Panevová et al., 2014)). Unfortunately, such a clear-cut criterion does not exclude all possible ambiguities. There are tricky contexts where the ambiguity could be removed only by a broader context or by the situation, see (25) and its two possible expansions in (26a) and (26b):

- (25) Vydala jsem výkřik, který kromě Artura musel slyšet kdekdo.
'I have given a scream which besides Arthur must have been heard by everybody.'

¹³For a detailed analysis of these constructions including other peculiarities occurring in Czech see (Panevová et al., 2014).

- (26) a. Vydala jsem výkřik, který kromě toho, že ho slyšel Artur, musel slyšet kdekdo.
'I have given a scream which in addition to that it was heard by Arthur must have been heard by everybody.'
- b. Vydala jsem výkřik, který kromě toho, že ho neslyšel Artur, musel slyšet kdekdo.
'I have given a scream which in addition to that it was not heard by Arthur must have been heard by everybody.'

The restructuring proposed for the type of sentences analyzed in this Section by means of an addition of the predicate corresponding to the governing predicate seems to be helpful from two points of view: One concerns the introduction of the means for splitting two functions conflated in the small clauses and the other is reflected in a more subtle classification of the list of adverbials adding an Addition and Exception as two new semantic units (functors) on tectogrammatical level.

3.4 Deletions in structures with comparison

Comparison structures are a very well known problem for any description pretending on restoration of elements missing in the surface shape to reach a complete representation of syntax and semantics of the sentences. In FGD two types of the comparison are distinguished: one is connected with the meaning of equivalence (introduced usually by the expression *jako* [as]; the subfunctor used in PDT has the label 'basic'), the other expresses the meaning of difference (it is introduced usually by the conjunction *než* [than]; the subfunctor used is called 'than'). There are some comparison structures where the restoration of elements missing on the surface seems to be easy enough from the point of view of semantics and from the point of view of the extent of the part inserted in the TR (see (27a), and its restored version (27b)).

- (27) a. Jan čte stejné knihy jako jeho kamarád.
'John reads the same books as his friend.'
- b. Jan čte stejné knihy jako (čte) jeho kamarád.
'John reads the same books as his friend (reads).'

Most comparisons are, unfortunately, more complicated, see the following examples and the arguments for the necessity of their extension:

- (28) a. Jan se choval na banketu jako v hospodě.
 'John behaved at the reception as in the pub.'
- b. Jan se choval na banketu (stejně), jako se (Jan) chová v hospodě.
 'John behaved at the reception (in the same way) as (John) behaves in the pub.'

In ex. (28a) we encounter a similar problem to the one we analyzed in Sect. 3.3. when discussing the modification of substitution, addition and of exception: in the comparison structure two semantic functions are conflated (comparison-basic and locative meaning in (28a)). Thus an artificial predicate sharing in this case the same value as the governing predicate (with the syntactic label comparison-basic) must be added into the extended representation. It serves as the head for the locative adverbial, too.

For many modifications of comparison, however, even a more complex reconstruction of comparison "small clauses" is needed. For an adequate interpretation of the surface shape of (29a) not only the shortened comparison structure with locative has to be expanded but also an "operator" indicating similarity of the compared objects is missing. For the identification of the similarity the expression as *stejný/stejně* [same/identically], *podobný/podobně* [similar/similarly] are used and this operator has to be added into the corresponding TR, see ex. (29b).

- (29) a. Požadavky jsou u Komerční banky jako u České spořitelny.
 'The requirements are at Commercial Bank as at Czech Saving Bank.'
- b. Požadavky jsou u Komerční banky (stejně) jako (jsou požadavky) u České spořitelny [#Some].
 Requirements are at Commercial Bank (same) as (are requirements) at Czech Saving Bank [#Some].

An adequate description of the type of comparison exemplified by ex. (29) (see Figure 2) requires

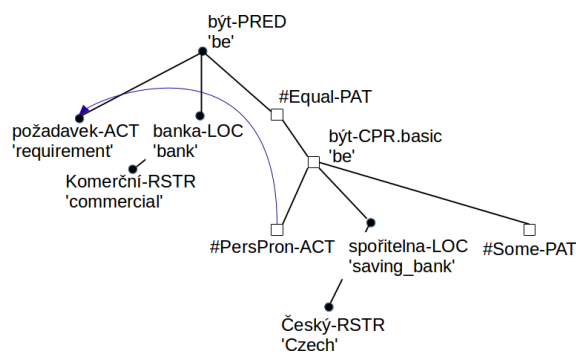


Figure 2: Deep structure of (29)

to add not only an artificial predicate the head of which copies the lemma of the main predicate, but also an operator indicating the type of comparison (#Equal, here with the meaning *stejný* [the same]). The artificial lemma #Some is used to stand for the lexically underspecified adjective/adverbial for both types of comparison, see (29b) and (30b).

While the extension of (29a) would be acceptable (at least semantically) in the form *Požadavky jsou u Komerční banky stejné jako (jsou stejné) u České spořitelny* [The requirements are at Commercial Bank the same as (are the same) at Czech Saving Bank], such type of extension is not acceptable with the comparison-than type (connected with the comparison of objects which are not similar), see (30). This sentence requires an artificial extension because the operators used for this type of comparison as *jiný/jinak* [different], *rozdílný* [different] have no semantic counterpart to be filled in the extended representation. The extension by the adjective *nějaký* [some] is given here by the fact that *jiný* has no single lexical counterpart for the expression of the Ministry situation in (30) (if the situation there is different, the appropriate adjective is actually unknown, it is underspecified).

- (30) a. Situace v armádě je jiná než na ministerstvu.
 'The situation in the army is different than at the Ministry.'
- b. Situace v armádě je jiná než (je situace) na ministerstvu [#Some].
 'The situation in the army is different than (the situation) at the Ministry is [#Some].

Our experience with the analysis of data in PDT indicates that the relations between the extension

of comparison modifications and the extent of their complete structure on the deep level differ very significantly, so that a more detailed classification would be useful.

4 Summary

We have analyzed four types of elided constructions in Czech and proposed their representation on the deep (tectogrammatical) level of syntactic description within a formal dependency-based description. From the point of view of the binary relation of the governor and its dependent, either the governor or the dependent may be missing and has to be reconstructed. A reconstruction of a dependent is e.g. the case of deletions connected with the pro-drop character of Czech ([*I came late*]), or in cases of a deleted general argument (*John sells at Bata [what][to whom]*), while a governor has to be reconstructed mostly in coordinated structures (*John likes Bach and Susan [likes] Beethoven; We know when [she came] and why she came*). In some types of deletions, the reconstruction concerns an introduction of a rather complex structure which is, however, needed for an appropriate semantic interpretation of the surface shape of the sentence, as illustrated by the comparison phrases and structures representing Addition and Exception. Our analysis focused on several types of the so-called systemic ellipsis, i.e. such that is given by grammatical rules or conditions or by a grammatically determined surface deletability; we have left aside textual ellipsis such as coordination, which is conditioned mostly by the context or by situation.

Surface deletions reflect the openness of the language systems to compress the information. However, for the description of meaning of such compressed structures more explicit means for an adequate and unambiguous description are needed.

Acknowledgments

The authors gratefully acknowledge the detailed remarks and suggestions by the three anonymous reviewers. We are deeply indebted to Barbora Hladká for her invaluable technical assistance. The work on this contribution was supported by the grant of the Czech Grant Agency P406/12/0557 and has been using language resources developed and/or stored and/or distributed by the LINDAT/CLARIN project of the Ministry of Education of the Czech Republic (project

LM2010013).

References

- Eduard Bejček, Eva Hajičová, Jan Hajič, Pavlína Jínová, Václava Kettnerová, Veronika Kolářová, Marie Mikulová, Jiří Mírovský, Anna Nedoluzhko, Jarmila Panevová, Lucie Poláková, Magda Ševčíková, Jan Štěpánek, and Šárka Zikánová. 2013. Prague Dependency Treebank 3.0. <https://ufal.mff.cuni.cz/pdt3.0>.
- Jan Hajič, Eva Hajičová, Marie Mikulová, Jiří Mírovský, Jarmila Panevová, and Daniel Zeman. 2015. Deletions and node reconstructions in a dependency-based multilevel annotation scheme. *Lecture Notes in Computer Science*, 9041:17–31.
- Bohuslav Havránek and Alois Jedlička. 1960. *Česká mluvnice [Czech Grammar]*. Praha:SPN.
- Richard A. Hudson. 1984. *Word Grammar*. Basil Blackwell Oxford [Oxfordshire] ; New York.
- Richard A. Hudson. 2007. *Language Networks: The New Word Grammar*. Oxford University Press.
- Aleksandr V. Isačenko. 1960. *Grammaticeskij stroj ruskogo jazyka v sopusstavlenii so slovackim*. SAV: Bratislava.
- Petr Karlík, Marek Nekula, and Zdenka Rusínová, editors. 1995. *Příruční mluvnice češtiny [Handbook of Grammar of Czech]*. Nakladatelství Lidové Noviny, Praha.
- Idan Landau. 2013. *Control in Generative Grammar*. Cambridge: Cambridge University Press.
- Igor Mel'čuk. 1988. *Dependency Syntax: Theory and Practice*. State University of New York Press.
- Marie Mikulová. 2011. *Významová reprezentace elipsy*. Studies in Computational and Theoretical Linguistics. Ústav formální a aplikované lingvistiky, Praha, Czechia.
- Timothy Osborne and Junying Liang. 2015. A survey of ellipsis in chinese. In *Proceedings of the Third International Conference on Dependency Linguistics, Depling 2015*, Uppsala, Sweden. Computational Linguistics group at Uppsala University in collaboration with Akademikonferens.
- Timothy Osborne, Michael Putnam, and Thomas Groß. 2012. Catena: Introducing a novel unit of syntactic analysis. *Syntax*, 15(4):354–396.
- Jarmila Panevová, Eva Hajičová, Václava Kettnerová, Markéta Lopatková, Marie Mikulová, and Magda Ševčíková. 2014. *Mluvnice současné češtiny 2, Syntax na základě anotovaného korpusu [Grammar of present-day Czech 2. Syntax of the basis of an annotated corpus]*, volume 2. Karolinum Praha, Prague.

- Jarmila Panevová. 1996. More remarks on control. In Eva Hajičová, Oldřich Leška, Petr Sgall, and Zdena Skoumalová, editors, *Prague Linguistic Circle Papers*, volume 2, pages 101–120. Amsterdam/Philadelphia: John Benjamins Publ. House.
- Adam Przepiórkowski and Alexandr Rosen. 2005. Czech and Polish raising/control with or without structure sharing. *Research in Language*, 3:33–66.
- Alexandr Rosen. 2006. O čem vypovídá pád doplňku infinitivu [What the case of the complement of the infinitive tells us]. In František Čermák and Renata Blatná, editors, *Korpusová lingvistika: Stav a modelové přístupy*, pages 254–284. Nakladatelství Lidové Noviny, Praha.
- Rudolf Růžička. 1999. *Control in Grammar and Pragmatics*. Amsterdam/Philadelphia: John Benjamins Publ. House.
- Petr Sgall, Eva Hajičová, and Jarmila Panevová. 1986. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Dordrecht:Reidel Publishing Company and Prague:Academia.