The 2014 Conference on Computational Linguistics and Speech Processing
ROCLING 2014, pp. 125-138

Semantic Representation of Ellipsis in the Prague Dependency Treebanks

Marie Mikulová*

Abstract

This article answers the question what is and what is not ellipsis and specifies criteria for identification of elliptical sentences. It reports on an analysis of types of ellipsis from the point of view of semantic representation of sentences. It does not deal with conditions and causes of the constitution of elliptical positions in sentences (when and why is it possible to omit something in a sentence) but it focuses exclusively on the identification of elliptical positions (if there is something omitted and what) and on their semantic representation in a treebank, specifically on their representation on the deep syntactic level of the Prague Dependency Treebanks. The theoretical frame of the approach to ellipsis presented in this article is dependency grammar.

Keywords: Ellipsis, Semantic Annotation, the Prague Dependency Treebanks.

1. Introduction

The analysis of well-formed sentences, some of whose constituents are missing, has been of central concern to computational linguists at least since the beginnings of the work in formal grammar. There has been a considerable amount of research on ellipsis from a variety of perspectives. Different approaches to an explanation of the procedures involved in assigning representations to sentences containing deletions have been developed (see, for example, Berman – Hestsvik, 1992 and Lappin – Benmamoun, 1999) but they have been mostly designed within the framework of constituency grammar.

The theoretical frame of our approach to ellipsis is dependency grammar. The dependency-based approach offers a totally different perspective on ellipsis. The constituency-based approach assigns more empty positions. The so-called gaps are assigned particularly when two constituents cannot be bracketed because they do not occur one next to the other in the surface form of a sentence. This discontinuity of two constituents does not impede the construction of a dependency tree. Only those gaps are perceived as ellipses when

* Charles University in Prague, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics, Czech Republic; E-mail: mikulova@ufal.mff.cuni.cz.

one of the constituents is not expressed at all in the surface form of the sentence. Thus, only a small part of the gaps (identified by constituency-based approaches) overlaps with the types of ellipses that have been defined by means of dependency syntax.

This article answers the question what is and what is not ellipsis in dependency grammar. It reports on an analysis of types of ellipsis from the point of view of semantic representation of sentences in a dependency treebank. We identify a boundary between the grammatical ellipses on the one hand and the accidental omissions on the other. We will distinguish morphological, surface and deep syntactic features of ellipses.

2. What is Ellipsis?

In attempting to characterize elided sentences in natural language we are faced with the problem of explaining how speakers or recipients are able to represent and interpret linguistic objects which, at least on the surface, are not present. We believe that one of the major questions is related to the reason why we suppose that something is missing in a sentence. There is an expectation of a certain lexical position to be realized, but this expectation is not fulfilled. However, who is the subject/bearer of the expectation? The speaker, the recipient, or the linguist? Such questions are also asked by Hlavsa (1981) who offers convincing reasons, saying that the subject that performs an analysis of ellipsis in a sentence can only be the linguist. Neither the speaker nor the recipient are aware of the presence of ellipsis (in terms of information that is being communicated). Sentences that are incomplete from the grammatical viewpoint (e.g. *Mary likes Bach, Susan Beethoven.*) still maintain their information value. Thus, the basic characteristics of ellipsis can be defined as incompleteness from the perspective of the grammatical system of the language. We do not claim that sentences cannot be analyzed from the viewpoint of their informational completeness. However, we refer to the types of elision that are supported by the grammatical system of the language and distinguish them from those that lack such a support (accidental omissions caused by a sudden interruption of a dialog for example).

If we say that ellipsis is a grammatical incompleteness, we have to specify the corresponding grammatical representation for each sentence. Therefore, we are going to analyze surface form of a sentence and search for a theoretical representation of this sentence in the grammatical system. Then it will be possible to give an unambiguous and reliable definition of ellipsis based on the relation between a particular surface form of a sentence and its description in the grammatical system. Our representation of a sentence is based on a solid, well-developed dependency syntax theory which is known as Functional Generative Description (for a detailed account of this framework, see e.g. Sgall et al. (1986)). This formal theoretical approach has already been applied to an analysis of multifarious linguistic phenomena, mostly concentrated on Czech but also in comparison with English, Russian or

some other languages. In the Functional Generative Description, the grammar has been described as a system of several layers. We work with three layers: morphological layer and two structural layers (surface and deep syntactic layer). Each layer has its own syntax. At each layer, the elementary units combine into complex units. In a simplified way, a complex unit of each layer consists of an autosemantic (lexical) base and function elements. At the syntactic layers, each complex unit of a sentence is classified as a governor or a dependent in relation to another one.

In our approach, ellipsis is an empty, non-expressed position in a sentence representation at the given grammatical layer (and what is beyond the boundary of the grammatical system, what is not defined in that system, it cannot be (grammatical) ellipsis; we treat accidental omissions, etc. here). Ellipsis is a relation; something is missing, because something else in the given sentence representation needs it for building complete complex unit or complete dependency structure. For example, at the morphological layer, a suffix or prefix cannot exist without a base word. Therefore, in the example *pre- and post-election discussion*, we analyze ellipsis of base word to the expressed prefix.

At the syntactic layers, we distinguish two pairs of ellipses:

A. Forming unit ellipses:

1. Ellipsis of an autosemantic base to a function element
2. Ellipsis of a function element to an autosemantic base

B. Dependency structure ellipses:

1. Ellipsis of a governing unit to a dependent (structural ellipses)
2. Ellipsis of (an obligatory) dependent to a governing unit (valency ellipses)

We define forming unit ellipses at both syntactic layers. At the surface syntactic layer, a complex unit consists of an autosemantic verb with its auxiliary verbs; prepositions are connected with nouns. So, we analyze ellipsis of an autosemantic verb to an auxiliary verb (type A.1) and ellipsis of an auxiliary verb to an autosemantic verb (type A.2); ellipsis of a noun to a preposition (type A.1) and ellipsis of a preposition to a noun (type A.2). As one deep complex unit, we treat idioms and connections of an autosemantic verb with a modal or phase verb.

The ellipsis of a governing unit to a dependent (type B.1) is the main type of ellipses at the surface syntactic layer. The ellipsis of an (obligatory) dependent to a governing unit (type B.2) is the main type of ellipses at the deep syntactic layer. Here, the ellipsis is a matter of valency (we work with the theory of valency that was treated particularly by Panevová (see Panenová, 1980; see also Urešová, 2011, about valency in the Prague Dependency Treebank).

3. Types of Ellipses

Ellipses are classified (as morphological, surface and deep syntactic) according to the layer of the grammatical system that has been assigned to the empty position. Table 1, 2 and 3 include a brief summary of the defined types of ellipsis with some simple examples that give a sufficient illustration of these types (the elided text is in square brackets).

Table 1. Morphological ellipses.

Word ellipses	
Ellipsis of end of word to prefix	*pre-*[election] *and post-election discussion*
Ellipsis of beginning of word to suffix	*in terms of the context(*[context]*s)*
Ellipsis of part of composite word	*two-*[seater] *or four-seater car*

Table 2. Surface syntactic ellipses.

Forming unit ellipses	
Ellipsis of auxiliary verb to autosemantic verb	*Peter will go to Prague and* [will] *visit his mother there.*
Ellipsis of autosemantic verb to auxiliary verb	*(Will you go?) Yes, I will* [go].
Ellipsis of noun to preposition	*in front of* [camera] *and behind the camera*
Ellipsis of preposition to noun	*in Prague and* [in] *Pilsen*
Ellipsis of verb to subordinating conjunction	*We know, when* [she came] *and why she came.*
Ellipsis of subordinating conjunction in dependent clause	*He said, that they would come and* [that they would] *stay the night.*
Ellipsis of second paratactic element	*Go away otherwise...* [?]
Structural ellipses	
Ellipsis of governing verb	*Mary likes Bach, Susan* [likes] *Beethoven.*
Ellipsis of governing noun	*Central* [Europe] *and Eastern Europe*

Table 3. Deep syntactic ellipses.

Forming unit ellipses	
Ellipsis of modal verb to autosemantic verb	*Peter wants to relax and* [wants to] *listen to music.*
Ellipsis of autosemantic verb to modal verb	*(Stay the night.) I could not* [stay].
Ellipsis of part of idiom	*They buried* [the hatchet] *and then dug up the hatchet again.*
Valency ellipses	
Textual ellipsis	*(Did the shop assistant pack the book?) Yes, he did* [what-the book].
General argument	*Jane sells at Bata* [what] [to whom].
Control	*The company planned* [who-the company] *to increase production.*
Reciprocity	*John a Mary met* [who-each other].

4. The Prague Dependency Treebanks

The Prague Dependency Treebank (PDT) is the first complex linguistically motivated treebank capturing also the deep syntactic structure of sentences. It is based on a dependency syntactic theory, on the Functional Generative Description. The treebank consists of continuous Czech texts mostly of the journalistic style analyzed at three layers of annotation: morphological (m-layer); surface syntactic (analytical, a-layer) and deep syntactic (tectogrammatical, t-layer). In addition to these three annotation layers there is also one non-annotation layer, representing the "raw-text". At this layer, called word layer (w-layer), the text is segmented into documents and paragraphs and individual tokens are recognized and associated with unique identifiers. At the m-layer each sentence is provided with morphological categories (lemma, tag). At the a-layer, a dependency tree captures surface syntactic relations such as Subject, Object, and Adverbial. The highest t-layer contains all the information that is encoded in the structure of the sentence and its lexical items. So, this layer captures the deep, semantico-syntactic structure, the functions of its parts, the "deep" grammatical information, coreference and topic-focus articulation including the deep word order. Figure 1 displays the relations between the neighboring layers as annotated and

represented in the data. Thus, for example, the Czech sentence *Byl by šel dolesa*, literally in English: 'He-was would went toforest.' contains past conditional of the verb *jít* (*byl by šel* 'he-was would went') and a typo (*dolesa* 'toforest').

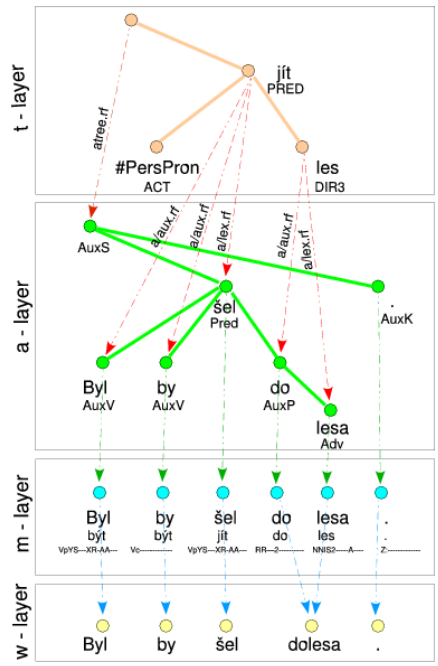

Figure 1. Linking the layers in PDT.

The total number of sentences annotated at all the three layers is 49,442, amounting to 833,357 (occurrences of) nodes. The PDT version 1.0 (with the annotation of the first two layers; Hajič et al., 2001) is available from the Linguistic Data Consortium, as is the version 2.0 (with the annotation of the third, deep syntactic layer; Hajič et al., 2006). The latest version PDT 3.0 (Bejček et al., 2013) with some additions (textual coreference, discourse relations, genre specification, multiword expressions) is available from the LINDAT/CLARIN repository. A similarly based annotation has been used for other Prague treebanks. The Prague Czech-English Dependency Treebank (Hajič et al., 2011) contains parallel PDT-like annotations of English texts (Wall Street Journal part of Penn Treebank) and of their professional translation to Czech. The Prague Dependency Treebank of Spoken Czech (it is planned to be released at the end of 2014) contains spontaneous dialogue speech, transcribed, reconstructed and further annotated in the PDT style.

5. Capturing Ellipsis Techniques

In the Prague Dependency Treebanks, ellipsis is treated at the highest t-layer. All types of ellipses given in Section 3 (excluding some minor exceptions) are captured at this layer. The principles of the build-up of the lower layers in PDT (namely that the number of nodes at these layers is identical to the number of tokens in the sentence) do not allow capturing ellipsis by an addition of a node (which is the most intuitive way of capturing ellipsis in corpora). Therefore, at the m-layer and a-layer, ellipsis is indicated only by a special attribute at such an expressed node that cannot be annotated according to usual rules because of an ellipsis in the sentence. At the t-layer, ellipses are represented by the linking the nodes of the highest t-layer with the nodes of the lower a-layer (Section 5.1) and by added nodes at the t-layer (Section 5.2).

5.1 Links of the t-layer to the a-layer

The relation between the t-layer and a-layer is rather complex. The linking between the layers is not only a technical question, but it carries a piece of linguistic information. It captures the transition from the (linguistic) meaning of a sentence to its form, transition from the deep syntactic structure to the surface expression. The basic principle of linking is as follows: there is a link from the t-node (node at the t-layer) to each a-node (node at the a-layer) that influences the value of some of its attribute. Based on the type of attribute that is influenced by the a-node two types of links to the a-layer are differentiated:

- **lex** = link to the a-node from which the t-node got its lexical meaning (or its biggest part). The a-node influences the value of the *lemma* attribute. Usually it is an a-node that represents an autosemantic word (noun, adjective, verb, adverb).
- **aux** = links to remaining a-nodes that influence values of attributes of the given t-node. Typically, the a-nodes affect values of syntactic and morphological attributes, such as *functor* (deep syntactic function), *subfunctor* (detailed classification of functors) and *gram* (the structure of attributes, the so-called grammatemes, which capture deep grammatical correlates of the morphological categories). These are usually a-nodes representing prepositions, conjunctions, auxiliary verbs, supporting expressions.

For example, in Figure 1, there are three links from the t-node representing the verb *jít* ‘to_go’: to the a-node representing word *šel* ‘went’, to the a-node representing word *by* ‘would’ and to the a-node representing the word *byl* ‘he-was’. The link to the first mentioned a-node (*šel* ‘went’) is *lex* type; the other two links are *aux* types.

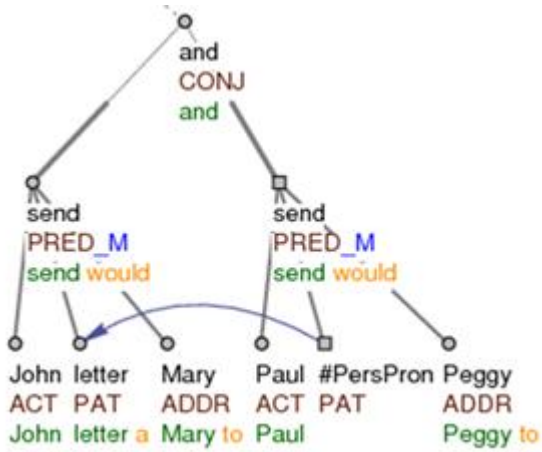

Figure 2. PDT-annotation: John would send a letter to Mary and Paul to Peggy.

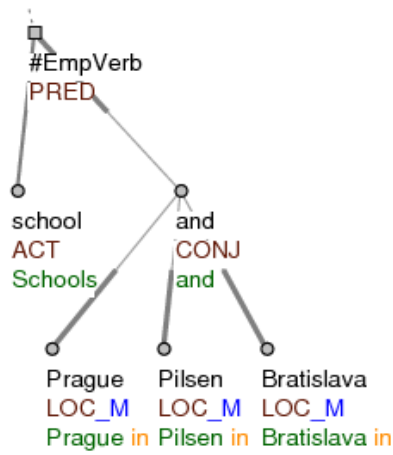

Figure 3. PDT-annotation: Schools in Prague, Pilsen and Bratislava.

There may be more links to one a-node – from various t-nodes. It happens in case of an ellipsis. For example, in the second clause of the sentence *John would send a letter to Mary, Paul to Peggy.*, a t-node will be added for the elided governing verb and there will be a lex-link from this added t-node to the a-node representing the word *send* and aux-link to the a-node representing the word *would*. The identical links will also be used in case of the t-node that represents a governing verb in the first clause. Thus there are two double links to a-nodes representing the words *send* and *would* from the t-layer and this is the way we capture the textual ellipsis of a governing unit (along with the addition of the t-node). In this case, the main indicator of ellipsis in the sentence is of course the added t-node. However, the textual ellipsis itself is captured by the double links to the a-layer (by the lex-link to the a-node from

the added t-node). See Figure 2; the linking a-nodes are written in green (lex-link) and orange (aux-link) color by each t-node.

Some types of ellipses are only captured via links to the a-layer. For example, the fact that prepositions in the fragment *Schools in Prague, Pilsen and Bratislava*. are not repeated, is represented only by an aux-link to the expressed preposition from each of the t-nodes that represent the names of cities (there is a link to the preposition *in* from three t-nodes; see Figure 3).

5.2 Addition of Nodes

In order to capture the whole meaning of a sentence, it is sometimes necessary to add t-nodes that lack their direct counterpart at surface form of sentence (at the a-layer). We distinguish two basic types of added t-nodes:

- **copy** = a t-node that has the same values of certain attributes (lemma, some grammatemes and valency frame) as an expressed t-node. We call this added t-node a "copy" of that t-node. There are double lex-links to the a-node that represents expressed copied word.
- **substitute** = an artificial t-node to which one of the following lemma substitutes has been assigned:

  a. *#EmpNoun*; *#EmpVerb*

  b. *#Cor*, *#Gen*, *#Oblfm*, *#PersPron*, *#Rcp*

Various types of ellipsis are captured by various types of added t-nodes. The textual ellipsis of governing verb or noun is captured using "copy" t-nodes. The system ellipsis (with no antecedent) of governing verb or noun is captured using "substitute" t-node with a lemma substitute mentioned in the list a) above. The various types of valency ellipses are captured using t-nodes with a lemma substitute mentioned in the list b). In case of textual ellipsis, there is a coreference (anaphora) arrow from the added t-node to the t-node representing expressed unit (anaphor).

In the tree, t-nodes representing expressed words are drawn as circles; added t-nodes are drawn as squares. In Figure 2, there is an example of capturing a textual ellipsis of the governing verb. In the second clause, there is an added t-node representing the governing verb and this t-node is a copy of the t-node representing the expressed verb in the first clause. The system ellipsis of the governing verb is captured in Figure 3. There is an added t-node with *#EmpVerb* lemma substitute on the governing verb position.

In Figure 2, there is also an example of textual valency ellipsis. The elided Patient *letter* (in the second clause) is captured using an added t-node with *#PersPron* lemma. There is a coreference arrow from the added t-node to the expressed Patient in the first clause. A similar example of textual valency ellipsis is also in Figure 4.

6. Summary of Representation of Ellipsis

This section brings an overview of the ways the defined ellipses (see the summary in Section 3) are represented at the t-layer. Structural ellipses (type B.1 in Section 2) are always captured by an addition of a "copy" or a "substitute" t-node (see Table 4; if the ellipsis is captured by a t-node with a lemma substitute, only the lemma is mentioned in the table). Valency ellipses (type B.2 in Section 2) are always captured by an addition of a t-node with a lemma substitute; the overview of valency types of ellipses and their lemmas substitute is presented in Table 5.

Table 4. Representation of structural ellipses in PDT.

	Textual ellipsis	**System ellipsis**
Ellipsis of governing verb	copy t-node lex-link	*#EmpVerb* no lex-link
Ellipsis of governing noun	copy t-node lex-link	*#EmpNoun* no lex-link

Table 5. Representation of valency ellipses in PDT.

Obligatory argument	*#PersPron* coreference arrow
Controlled argument	*#Cor* coreference arrow
Reciprocity	*#Rcp* coreference arrow
General argument	*#Gen* no coreference arrow

At the t-layer, we also capture both types forming unit ellipses: surface syntactic (forming unit ellipses in Table 2) and deep syntactic (forming unit ellipses in Table 3). The method of capturing is similar for both types. The ellipsis of an autosemantic base to a function element (type A.1 in Section 2) is captured according to the same rules that apply to the ellipsis of a governing verb or noun, i.e. with the help of the "copy" t-node (in case of textual ellipsis) or "substitute" t-node with the *#EmpVerb* or *#EmpNoun* lemma (in case of system ellipsis). There are aux-links to a-nodes representing expressed function elements. For a clear overview, see Table 6. An example of capturing an ellipsis of an autosemantic base to a function element is in Figure 4. There is an ellipsis of the verb *obtain* to the modal verb *want* in the second clause.

Ellipses of a function element to an autosemantic base (type A.2 in Section 2) are always captured only by means of links to the a-layer. See the example *Schools in Prague, Pilsen and Bratislava* in Figure 3 where ellipsis of a preposition to nouns is captured.

Table 6. Ellipsis of autosemantic base to function element in PDT.

Textual ellipsis	System ellipsis
copy t-node	*#EmpVerb* / *#EmpNoun*
lex-link	no lex-link
aux-links	aux-links

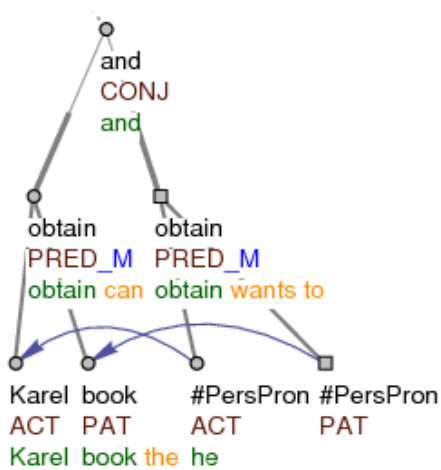

Figure 4. PDT-annotation: Karel can obtain the book and he wants to.

The majority of the defined types of ellipses (see their summary in Section 3) is captured in the PDT. However, the morphological word ellipsis (see Table 1) still remains unannotated. We suggest that special morphological tags should be implemented to resolve the cases when only a part of a word (prefix, suffix, ending, part of a composite word) is separated as an individual lexical unit based on the principle of tokenization "from a space to a space". Next, we suggest capturing the word ellipses in a similar way to the representation of ellipsis of an autosemantic base to a function element (see Table 6). The lemma of the t-node representing the elided word should get a non-elided, full lexical form and the lexical and morphological attributes of this t-node should correspond to a non-elided reconstructed form. There should be a lex-link to a-node representing a non-elided word (in case of textual ellipsis) or there should be no link at all (in case of system ellipsis). The a-node representing the expressed part of a word (prefix, suffix, ending, part of a composite word) should be captured as an aux-linking a-node.

7. Conclusion

In the article, we delimitate the theoretical basis of ellipsis, bring a productive classification and describe its semantic representation in the Prague Dependency Treebanks.

The Prague Dependency Treebanks has already been used in many projects, both in linguistics research and for the development of NLP learning algorithms and software tools. The analytical and later the tectogrammatical layer has been used as a basis for conversion to the well-known CoNLL format in the 2007 and 2009 shared tasks (Hajič et al., 2009). In these shared tasks, over 30 parsers and SRL tools have been created, as described in the proceedings of CoNLL 2007 and CoNLL 2009 Shared Tasks. Parallel Czech-English Dependency Treebank has been also used for learning parsers and sematic role labelers, which have in turn been used for machine translation projects (Bojar et al., 2009; Tamchyna et al., 2014; Mareček, Popel, Žabokrtský, 2010). Based on the data and experience with the PDT, a new multilingual treebank HamleDT with 30 treebanks in different languages have been converted to a uniform format and made available (Rosa et al., 2014).

Acknowledgments

The research reported in the article was supported by the Ministry of Education, Youth and Sport of Czech Republic within the project LINDAT-Clarin LM2010013. The participation on ROCLING 2014 conference was also supported by the Škoda Auto company.

Reference

Bejček, E. et al. (2013). *Prague Dependency Treebank 3.0*. Data/software, Institute of Formal and Applied Linguistics, Charles University in Prague, Prague. URL: http://ufal.mff.cuni.cz/pdt3.0/

Berman, S. – Hestvik, A. (eds.) (1992). *Proceedings of the Stuttgart Ellipsis Workshop*. Arbeitspapiere des Sonderforschungsbereichs 340, Bericht No. 29, University of Stuttgart, Stuttgart.

Bojar O., Callison-Burch Ch., Hajič J., Koehn P. (2009): Special Issue on Open Source Machine Translation Tools, *The Prague Bulletin of Mathematical Linguistics*, Charles University in Prague, Prague, No. 91, ISBN 978-80-904175-1-9, ISSN 0032-6585.

Böhmová, A., Hajič, J., Hajičová, E., Hladká, B. (2003). The Prague dependency treebank: a 3-level annotation scenario. In *Treebanks: Building and using parsed corpora*, ed. A. Abeillé, Kluwer, Dordrecht, 103-128.

Hajič J, Ciaramita M., Johansson R., Kawahara D., Martí M. A., Marquez L., Meyers A., Nivre J., Padó S., Štěpánek J., Straňák P., Surdeanu M., Xue Nianwen, Zhang Yi (2009): The CoNLL-2009 Shared Task: Syntactic and Semantic Dependencies in Multiple Languages. In: *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL): Shared Task*, Association for Computational Linguistics, Boulder, CO, USA, ISBN 978-1-932432-29-9, 1-18.

Hajič, J. et al. (2006). *The Prague Dependency Treebank 2.0*. CD-ROM. Cat. No. LDC2006T01, Linguistics Data Consortium, Philadelphia.

Hajič, J., Hajičová, E., Panevová, J. et al. (2012). Announcing Prague Czech-English Dependency Treebank 2.0. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, European Language Resources Association, Istanbul, ISBN 978-2-9517408-7-7, 3153-3160.

Hajič, J. et al. (2011). *Prague Czech-English Dependency Treebank 2.0*. Data/software, Institute of Formal and Applied Linguistics, Charles University in Prague, Prague.

Hajič, J. et al. (2001). *Prague Dependency Treebank 1.0 (Final Production Label)*. CD-ROM. Cat. No. LDC2001T10, Linguistic Data Consortium, Philadelphia, ISBN 1-58563-212-0.

Hlavsa, Z. (1981). On Syntactical Ellipsis. In *Satzsemantische Komponenten und Relationem im Text*, eds. Daneš, F., Viehweger, D., ÚJČ ČSAV, Prague, 119-128.

Lappin, S. – Benmamoun, E. (eds.) (1999). *Fragments. Studies in Ellipsis and Gapping*. Oxford University Press, New York, ISBN 0-19-512302-6.

Mareček D., Popel M., Žabokrtský Z. (2010). Maximum Entropy Translation Model in Dependency-Based MT Framework. In: *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR*, Association for Computational Linguistics, Uppsala, Sweden, ISBN 978-1-932432-71-8, 201-201.

Mikulová, M. et al. (2006): *Annotation on the tectogrammatical level in the Prague Dependency Treebank. Annotation manual*. Technical report no. 2006/30, Institute of Formal and Applied Linguistics, Charles University in Prague, Prague.

Marcus, M. et al. (1995). *Treebank 2*. CD-ROM. Cat. No. LDC95T7, Linguistics Data Consortium, Philadelphia.

Rosa R., Mašek J., Mareček D., Popel M., Zeman D., Žabokrtský Z. (2014). HamleDT 2.0: Thirty Dependency Treebanks Stanfordized. In: *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*, European Language Resources Association, Reykjavík, Iceland, ISBN 978-2-9517408-8-4, 2334-2341.

Sgall, P. et al. (1986). *The Meanings of the Sentence in Its Semantic and Pragmatic Aspects*. Dordrecht, D. Reidel Publishing Company - Praha, Academia, Prague.

Tamchyna A., Popel M., Rosa R., Bojar O.: CUNI in WMT14 (2014). Chimera Still Awaits Bellerophon. In: *Proceedings of the Ninth Workshop on Statistical Machine Translation*, Association for Computational Linguistics, Baltimore, MD, USA, ISBN 978-1-941643-17-4, 195-200.

Urešová Z. (2011): *Valenční slovník Pražského závislostního korpusu (PDT-Vallex)*. Institute of Formal and Applied Linguistics, Charles University in Prague, Prague., ISBN 978-80-904571-1-9.