Automatic Processing of Linguistic Data as a Feedback for Linguistic Theory

Vladislav Kuboň, Markéta Lopatková, and Jiří Mírovský

Charles University in Prague
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics
{vk,lopatkova,mirovsky}@ufal.mff.cuni.cz

Abstract. The paper describes a method of identifying a set of interesting constructions in a syntactically annotated corpus of Czech – the Prague Dependency Treebank – by application of an automatic procedure of analysis by reduction to the trees in the treebank. The procedure clearly reveals certain linguistic phenomena that go beyond 'dependency nature' (and thus generally pose a problem for dependency-based formalisms). Moreover, it provides a feedback indicating that the annotation of a particular phenomenon might be inconsistent.

The paper contains discussion and analysis of individual phenomena, as well as the quantification of results of the automatic procedure on a subset of the treebank. The results show that a vast majority of sentences from the subset used in these experiments can be analyzed automatically and it confirms that most of the problematic phenomena belong to the language periphery.

1 Introduction

Gathering various kinds of linguistic resources has become one of the major activities of many linguists during the past twenty years.

One of the factors which substantially influence the quality of linguistic resources, is the annotation consistency. The annotation consistency is very difficult to maintain especially for large data resources. In the process of annotation, the annotators have to make a huge number of small decisions related especially to borderline phenomena or to phenomena from the language periphery. A series of such decisions may lead to an annotation which, when viewed from the outside, may look unnatural or inconsistent.

In this paper we try to tackle the problem of annotation correctness and consistency of a large scale linguistic resource with very detailed syntactic annotation. This resource, the Prague Dependency Treebank (PDT) [1], actually provided a feedback for the theory it has been based upon, namely the Functional Generative Description [2]. This interesting fact has been described esp. in [3]. Our investigation takes this problem one step further – it identifies phenomena which have been annotated in a problematic way. It provides the feedback to the annotation, not to the underlying theory directly (although some of the findings may have consequences for the theory itself).

F. Castro, A. Gelbukh, and M. González (Eds.): MICAI 2013, Part I, LNAI 8265, pp. 252–264, 2013.

In our experiments we concentrate on the crucial relation for all dependency-based theories, the relation of dependency itself. We are going to define it through analysis by reduction (introduced in Section 2), a stepwise simplification of a sentence preserving its correctness [4, 5]. We want to gain better insight into the problem by means of the application of a semi-automatic procedure (requiring, of course, a subsequent manual checking) on a relatively large subset of PDT data. The results are presented in Section 4. In this way we can verify the concept against real data and, at the same time, shed more light on the way how individual linguistic phenomena are annotated.

1.1 The Background

In the world of dependency representation, there are three essential (and substantially different) syntactic relationships, namely 1. *dependencies* (the relationship between a governing and a modifying sentence member, as e.g. a verb and its object, or a noun and its attribute), 2. *'multiplication'* of two or more sentence members or clauses (esp. coordination), and 3. *word order* (i.e., the linear sequence of words in a sentence).[1] In this paper we are concentrating on the phenomenon 1. and 3, i.e., on the relationships of dependency and word order. These two basic syntactic relationships (dependency and word order) are relatively complex especially in languages with a higher degree of word order freedom.

Within dependency linguistics, these relationships have been previously studied especially within the Meaning-Text Theory: the approaches aiming at the determination of dependency relations and their formal description are summed up esp. in [7]. An alternative formal description of dependency syntax can be found in [8]. Our approach is based on the Czech linguistic tradition represented mainly in [2].

The second notion important for our experiments is the notion of an *analysis by reduction*. This notion helps to define the dependency relations: if, in the course of a stepwise reduction of a sentence, one of the words creating possible governor-modifier pair can be deleted without changing the distribution properties of the pair (i.e., the ability to appear in the same syntactic context) then it is considered as a modifying one (dependent on the latter one). This is applicable to endocentric constructions (as, e.g. *small table, Go home!*); for exocentric constructions (as *Peter met Mary*), the principle of analogy on the part-of-speech level is applied. Roughly speaking, as a sentence containing an intransitive verb is correct (*e.g., Peter sleeps*), an object is considered as dependent on a (transitive) verb as well; similarly for other types of objects as well as for a subject (as subject-less verbs exist in Czech, e.g., *Prší* '(It) rains'), see [2, 4].

The reason for exploiting the analysis by reduction is obvious: it allows for examining dependencies and word order independently. The method of AR has been described in detail in [4, 5], its formal modeling by means of restarting automata can be found in [9–11]. A brief description of its basic principles follows in Section 2.

[1] [6] considers linear order vs. structural order and also divides the structural relationships between connexion (now dependency) and junction (coordination).

2 Methodology – Analysis by Reduction

Let us now describe the main ideas behind the method used for sentence analysis. *Analysis by reduction (AR)* is based on a stepwise simplification of an analyzed sentence. It defines possible sequences of reductions (deletions) in the sentence – each step of AR is represented by *deleting* at least one word of the input sentence; in specific cases, deleting is accompanied by a *shift* of a word form to another word order position.

Let us stress the basic constraints imposed on the analysis by reduction, namely:

(i) the obvious constraint on preserving individual word forms, their morphological characteristics and/or their surface dependency relations;
(ii) the constraint on preserving the correctness (a grammatically correct sentence must remain correct after its simplification);
(iii) the application of the shift operation is limited to cases where it is enforced by the correctness preserving principle of AR.

Note that the possible order(s) of reductions reflect dependency relations between individual sentence members, as it is described in [5, 11]. The basic principles of AR can be illustrated on the following Czech sentence (1).

Example

(1) *Marie se rozhodla nepřijít.*
Marie – *refl* – decided – not to come
'Marie has decided not to come.'

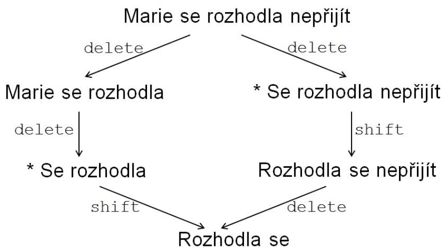

Fig. 1. Scheme of AR for sentence (1)

The analysis by reduction can be summarized in the scheme in Fig. 1. The sentence can be simplified in two ways:

(i) Either by simple deletion of the dependent infinitive verb *nepřijít* 'not to come' (see the left branch of the scheme).
(ii) Or by deleting the subject *Marie*[2] (the right part of the scheme). However, this simplification results in an incorrect word order variant starting with a clitic

[2] Note that Czech is a pro-drop (null-subject) language; thus it is possible to reduce a sentence subject (if present at all) at any moment and the sentence remains correct (if some word order constraint is not violated).

se[3] $\rightarrow_{shift}$ **Se rozhodla nepřijít.* Thus the word order has to be adjusted (by applying a shift) in order to preserve the syntactic correctness of the sentence: $\rightarrow_{shift}$ *Rozhodla se nepřijít.* '(She) decided not to come'.

Now, we can proceed in a similar way until we get the minimal correct simplified sentence *Rozhodla se.* 'She decided.'

We can notice that the order of reductions reflects the dependency relations in the corresponding dependency tree. Informally, the words are 'cut from the bottom of the tree'; i.e., a governing node must be preserved in a simplified sentence until all its dependent words are deleted, see [4]. In other words, AR corresponds to the dependency tree for sentence (1).

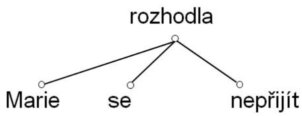

Fig. 2. Dependency tree of sentence (1); PDT-like (surface) syntactic annotation

Projectivity. The phenomenon of (non-)projectivity is one of very interesting and problematic language phenomena [14].[4] As a supplementary result, we are going to quantify how many sentences can be completely processed by a simple analysis by reduction (analysis by reduction with delete operation only). For this reason, we allow only for projective reductions. In other words, dependent word in a distant position cannot be deleted (with the only exception of limited technical non-projectivities caused, e.g., by prepositions).

The constraint allowing only projective reductions makes it possible to describe a core projective word order. It shows that – even within projective constructions – certain constraints on word order exist, esp. in connection with the position of clitics.

Let us demonstrate the processing of non-projective reductions on the following example (2) (based on [15], modified).

Example

(2) *Petr se Marii rozhodl tu knihu nekoupit.*
Petr – *refl* – to Mary – decided – the book – not to buy
'Petr has decided not to buy the book to Mary.'

The word *Marii* 'Mary' (indirect object of the verb *nekoupit* 'not to buy') cannot be reduced because it is 'separated' from its governing verb by the main predicate

[3] Czech has strict grammatical rules for clitics – roughly speaking, they are usually located on the sentence second (Wackernagel's) position, see esp. [12, 13]

[4] Informally, projective constructions meet the following constraint: having two words n_{gov} and n_{dep}, the second one being dependent on the first one – then all words between these two words must also (transitively) depend on n_{gov}.

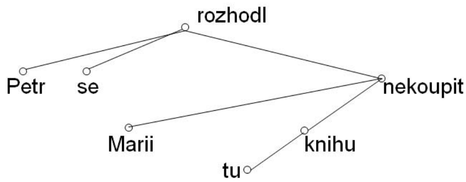

Fig. 3. Dependency tree of sentence (2)

rozhodl 'decided' (i.e., by the root of the dependency tree) and thus the relation *Marii – nekoupit* 'to Mary – not to buy' is represented by a non-projective edge in the dependency tree. Thus within the projective AR, a shift must be performed to make the reduction possible: *Petr se Marii rozhodl tu knihu nekoupit.* $\rightarrow_{shift}$ *Petr se rozhodl Marii tu knihu nekoupit.* $\rightarrow_{delete}$ *Petr se rozhodl tu knihu nekoupit.*

3 Semi-automatic Application of AR on the PDT Data

3.1 Data

For humans, especially for native speakers of a particular natural language, it is easy to apply the analysis by reduction, at least when simple sentences are concerned. However, this application exploits the fact that the human understands the sentence and that (s)he is naturally able to reduce the sentence step by step. When we are aiming at applying AR automatically, we have to 'substitute' (at least to some extent) the understanding using the syntactically annotated data (with subsequent manual correctness checking).

For our experiments we make use of the data from the Prague Dependency Treebank 2.0 (PDT, see [1]).[5] The syntactic structure – given by dependency trees (a single tree for a single sentence) – actually guided the process of AR.

The PDT contains very detailed annotation of almost 49,500 Czech sentences. The annotation is performed at multiple layers, out of which the analytical layer – describing (surface) syntactic structure employing so called analytical functions – is the most relevant for our experiments; we are taking into account only training data (38,727 sentences) (leaving the test set for evaluation in the future).

Investigating individual linguistic phenomena is easier if only simple sentences are taken into account. In the initial phase of our experiments, we concentrate on sentences which do not contain phenomena of obviously non-dependent character (esp. coordination, apposition, and parentheses). We also focus only on sentences with a single *finite verb* (and thus typically consisting of a single clause only). Note that even these sentences can have quite complex structure, including non-projectivities, see ex. (2).

For obtaining a suitable set of test sentences for AR as well as for searching the data, we exploit a PML-TQ search tool, which has been primarily designed

[5] `http://ufal.mff.cuni.cz/pdt2.0/`

for processing the PDT data. PML-TQ is a query language and search engine designed for querying annotated linguistic data [16], based on the TrEd toolkit [17]. TrEd with the PML-TQ extension allows users to formulate complex queries on richly annotated linguistic data.

This tool makes it possible to extract a subset of the corpus containing sentences with desired properties (we want to filter out sentences with too many phenomena), namely the sentence length limited to 10-25 tokens; no coordination and apposition nodes; no parentheses; just one finite verb; and no numerals.

Out of the 38,727 sentences of the training data of PDT, only 2,453 sentences remained after the application of this preprocessing filter. Although this number constitutes only 6.33% of the training set, it is still too big for manual inspection and it clearly shows the necessity of a semi-automatic method of applying AR to the data.

3.2 The Automatic Procedure

The automatization of the analysis by reduction requires a very careful approach. It is necessary to guarantee the correctness of the analyzed sentences in each step of the AR. The process is oriented bottom-up, it starts with the leaves of the dependency tree and it removes all dependent nodes stepwise, preserving one very important word-order condition, namely the condition that the neighboring nodes must always be removed first, followed by those which are connected by projective edges. The second very important condition is the *preservation of non-projectivity.* A node cannot be reduced if this reduction would result in some non-projective edge becoming projective, see sentence (2).

Let us now describe how individual linguistic phenomena are handled by the automatized AR.

Arguments, Adjuncts and Attributes. These categories are actually the simplest ones because they are the most regular ones. This group includes all nodes marked by analytical functions for attributes (Atr), adverbials (Adv), objects (Obj) and subjects (Sb). All these types of nodes can be reduced automatically, they represent the most natural dependency relationships.

Prepositions. This part of speech actually represents one of the most obvious examples of a phenomenon which is not naturally of a dependency nature. Prepositions typically serve as a kind of a morpho-syntactic feature of a noun (similarly as, e.g. a morphological case). However, many dependency-based theories and treebanks (including PDT) prefer to represent prepositions as governing nodes for the whole prepositional groups.

For our procedure it means that if a node is governed by a preposition (analytical function AuxP), it is necessary to reduce both nodes at once, in a single step. This has an important consequence: prepositions (AuxP) are also ignored when projectivity is tested – i.e., if the only source of a non-projective edge is a preposition, the sentence is treated as projective (this is justified by rather technical annotation of prepositions in PDT).

A special category of multiword prepositions is handled in a similar way – all words depending on the preposition (with the AuxP analytical function) are being ignored by AR until the governing preposition is being reduced – at this moment all the dependent words of the preposition are being reduced as well. For example, in the sentence *Byl zrušen s výjimkou programu.* 'It was cancelled with the exception of the program.' it is necessary to delete both the noun *programu* 'program' and the multiword preposition *s výjimkou* 'with (the) exception of' in one step. The sentence will then be reduced to *Byl zrušen.* '(It) was cancelled.'

Comparison. The constructions *čím – tím* (as in, e.g., *čím starší, tím lepší* 'the older the better') constitutes a very special case, which goes beyond the dependency nature of AR. For the time being, we skip sentences with this construction in our experiments.

Other types of comparisons, as, e.g., *než* 'than' (analytical function AuxC), *jako* 'as' (AuxY or AuxC combined with ellipsis marked ExD), do not cause any problems, they are always reduced together with their last child. Let us demonstrate this reduction on the sentence

Example

(3) *Míra nezaměstnanosti by se měla vyvíjet protikladně,*
rate – unemployment – *cond* – *refl* – should – develop – opposite –
než ve standardní ekonomice.
than – in – standard – economy
'The unemployment rate should develop in the opposite manner than in a standard economy.'

After deleting the adjective *standardní* 'standard', the stepwise reduction deletes within a single step of AR:
- the noun *ekonomice* 'economy' (analytical function ExD)
- together with the governing preposition *ve* 'in' (AuxP), and
- the comma 'AuxX' (a node depending on *než* 'than' (AuxC) node;
- further, the conjunction *než* 'than' (AuxC) is reduced.

Thus the reduced sentence *Míra nezaměstnanosti by se měla vyvíjet protikladně.* 'The unemployment rate should develop in the opposite manner.' is obtained. The reduction process may then continue further.

Clitics. Clitics have a relatively strictly defined position in grammatical Czech sentences – they must typically be located on a sentence second (Wackernagel's position) and thus they constitute a serious obstacle for our automatic procedure – they are reduced only together with their governing word, and, on top of that, no reduction may be performed which would leave a clitic on the sentence first position and thus make the reduced sentence ungrammatical.

Let us use the partially reduced sentence from the previous subsection as an example. It can be further reduced in several steps into → *Míra by se měla.* 'The rate should.' and no further, because reducing the subject *Míra* 'rate' would leave the clitics *by* and *se* in the sentence first position.

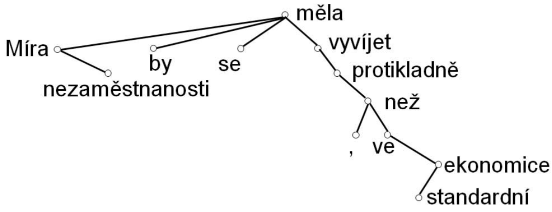

Fig. 4. Dependency tree of sentence (3)

Particles. Particles (AuxY) are in principle being reduced in a standard way (similarly as e.g. adverbials), it is only necessary to make sure that their reduction will not result in word order constraint violation. However, there is a special set of particles which constitute an exception – *coby, jako, jakoby, jakožto* 'as, like, as if' are being reduced together with their parent, similarly as in the case of comparison.

Emphasizing Expressions. If the word order permits it, emphasizing particles (AuxZ) can be reduced in the same way as, e.g., adverbials. If a prepositional group is involved in the emphasizing expression, it is reduced as a single unit. When checking the word order constraint, the nodes marked by AuxY (particles) are being ignored.

Punctuation and Graphical Symbols. Reduction of these symbols (AuxX, AuxG) can be applied when the governing word is being reduced, if the word order constraint permits it. Some problematic issues are caused by inconsistent treatment of expressions containing dashes or apostrophes (as, e.g., names like *Marie-Anne, Koh-i-noor, B-konto, Preud'homme* etc.) – these expressions clearly constitute a single unit, but they are not understood as such in some cases at the analytical layer of the treebank.

Conjunctions. A (subordinating) conjunction (AuxC) is reduced together with its last daughter node (which is not an emphasizing word (AuxZ), graphical symbol (AuxG) or punctuation itself (AuxX); if present, all these nodes are reduced, too). This simple rule may be used also due to the fact that sentences containing coordinations were left out from our sample set of sentences, their reduction may be more complicated when sentences with coordinations are included into the sample set in the future.

Full Stop. Sentence ending punctuation (AuxK) is always reduced as a final step of AR.

Note that in some cases, we do not insist on a complete reduction (with only the predicate left at the end). Even with the set of test sentences mentioned above and the uncomplete reductions, the automatic AR gives us interesting results – see the tables in the following section. Apart from the numerical results, this approach also helped to identify some annotation inconsistencies (for example the four particles listed above or the annotation of names containing special characters) or phenomena which do not have dependency nature and their annotation thus may cause technical problems (prepositions, some types of comparison, etc.).

4 Analysis of the Results of the Automatic Procedure

4.1 Quantitative Analysis of the Results

Let us now quantify and analyze the results of the automatic AR applied on the test sentences from the PDT. First of all, Table 1 provides numbers of sentences where specific problematic phenomena appear (from the complete set of the training data from PDT, i.e., from 38,727 sentences).

Table 1. Numbers of sentences with specific syntactic phenomena in PDT

	phenomenon
12,345	sentences containing clitic(s); out of which 3,244 non-projective
850	with the comparison or complement (AuxY or AuxC) introduced by *coby, jako, jakoby, jakožto*; out of which 451 non-projective
895	with the comparison expressed by *než* (AuxC); out of which 323 non-projective
844	with the comparison with ellipsis (ExD); out of which 302 non-projective
32	with the comparison expressed by *čím – tím*; out of which 17 non-projective

Let us mention the reasons why we consider these phenomena problematic from the point of view of AR. First, clitics have a strictly specified position in a Czech sentence; thus they may cause a complex word order (including number of non-projective edges, see examples (1) and (2)). Second, a comparison (frequently accompanied by ellipses) has also complex and non-dependency character, as shown in example (3).

Let us now look at the results of simple (projective) reductions as described in the previous sections. The first column of Table 2 describes the number of nodes (= word forms) to which sentences were reduced; the second column gives the number of corresponding sentences and the third column gives their proportion with respect to the whole test set of 2,453 sentences.

We can see that our 'careful' automatic model of the simple AR (projective AR without shifts) can process almost 67% of the test set (plus 15.6% sentences are reduced into simple structures with 2 or 3 nodes). Note that (out of 2,453 test

Table 2. PDT – number of resulting nodes for analyzed sentences

nodes	sentences	%	cumulative coverage
1	1,640	66.86	
2	29	1.18	68.04
3	354	14.43	82.47
4	235	9.58	92.05
5	113	4.61	96.66
6	44	1.79	98.45
7	21	0.86	99.31
8	10	0.41	99.72
9	5	0.20	99.92
10	2	0.08	100.00

sentences), 282 sentences were non-projective (and thus cannot be fully reduced in the course of projective AR).

The results presented in Table 2 actually support the claim that the automatic procedure works surprisingly well given the complexity of the task. It is able to reduce more than 92% of input sentences to trees with 4 or less nodes. On top of that, it fails to reduce the tree (by a failure we understand the reduction to 7 or more nodes) in 1.6% of cases only.

4.2 Manual Analysis of the Results

After a manual analysis of the sentences that were reduced automatically to two nodes (29 in total), we can see that 23 sentences contain a clitic (dependent on the predicate), which prevents the full reduction; or an auxiliary verb (6 cases); or punctuation (1 case) (both auxiliary verbs and punctuation are represented as separate nodes in PDT). Further, 5 sentences which start with subordinating conjunction complete the list (as, e.g., → *Že rozeznáte* ‘That (you) recognize’). The results for sentences that were reduced to 3 and 4 nodes, respectively, are shown in Table 3.

Table 3. PDT – the analysis of sentences reduced to 2, 3 and 4 nodes

resulting in 1 node	resulting in 2 nodes	phenomenon	resulting in 3 nodes	resulting in 4 nodes
1,640	29	# sentences	354	235
–	23 / 0	1 clitic / 2 clitics	307 / 3	149 / 10
0	6 / 1	aux. verb / punctuation	74 / 0	30 / 2
–	–	1 non-proj. / 2 non-proj.	37 / –	82 / 2
0	5	others	0	0

In order to illustrate the most complicated cases, let us look at one sentence from the 'bottom' part of Table 2.

Example

(4) *V ČR se využívá v menším rozsahu než v zemích.*
In CR– *refl* – is used – in – smaller – extent – than – in countries
'In the Czech Republic it is used in a smaller extent than in countries.'

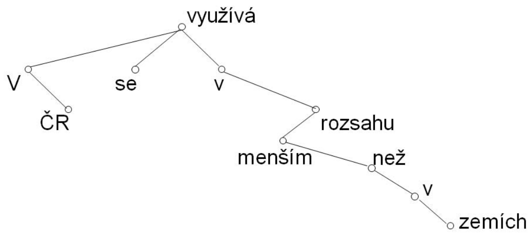

Fig. 5. Dependency tree for sentence (4)

In this case, 10 nodes remain as a result of the simple AR: The prepositional group *V ČR* 'In the Czech Republic' must be preserved in order to preserve correct position of the reflexive particle (clitic) *se*; further, the non-projective edge *menším–než* 'smaller–than' in the comparison ('separated' by the governing node *rozsahu* 'extent') stops the process of AR. So, as we can see, the sentence contains a complicated interplay of three phenomena – nonprojectivity, clitic and comparison, and thus constitutes a substantial challenge to the AR.

Let us now summarize the results from the point of view of language phenomena. Our experiments with the automatic AR have revealed that for the phenomena belonging to the language core (as, e.g., arguments, adjuncts and attributes) the AR works very smoothly (at this point it is important to remind that coordinations, appositions and parentheses were filtered out for this phase of our experiments). A vast majority of problems or irregularities is caused by the phenomena from language periphery (where, in many cases, the relations are not dependency-based and the syntactic structure is unclear).

The experiments also gave a feedback to the annotation of PDT – it has revealed that certain particles (with the analytical function of AuxY) which introduce comparison (lemmas *coby, jako, jakoby, jakožto* 'as, like, as if') behave in a completely different way than other nodes with the same analytical function. From the point of view of AR they have similar properties as the conjunctions *než* 'than', *jako* 'as' (which are annotated as AuxC). It is not clear why they have been annotated in this way and whether there are any reasons for this annotation, especially when the annotation manual of PDT does not give any answer to this question [18].

Conclusion and Perspectives

In this paper we have tried to achieve a deeper insight into the phenomenon of dependency and its annotation. The investigation has been performed by means of a semi-automatic analysis of a subset of a large corpus. This analysis proved the consistency of the annotation of the majority of sentences from the selected subset of PDT on the one hand, on the other hand it also helped to identify problematic constructions and syntactic phenomena.

In the future we would like to continue the research by examining more complex sentences containing linguistic phenomena that have been left out in this initial experiment. The most natural phenomenon which causes serious problems to all dependency theories and which definitely requires further investigation, is coordination. It would also be interesting to develop a (semi-)automatic method for an optimal application of the shift operation that would allow for projectivization of processed sentences and their full reduction.

Acknowledgments. The research reported in this paper has been supported by the Czech Science Foundation GA CR, grant No. GA P202/10/1333 and partially grant No. GA P406/2010/0875. This work has been using language resources stored and/or distributed by the LINDAT-Clarin project of MŠMT (project LM2010013).

References

1. Hajič, J., Panevová, J., Hajičová, E., Sgall, P., Pajas, P., Štěpánek, J., Havelka, J., Mikulová, M., Žabokrtský, Z., Ševčíková-Razímová, M.: Prague Dependency Treebank 2.0. LDC, Philadelphia (2006)
2. Sgall, P., Hajičová, E., Panevová, J.: The Meaning of the Sentence in Its Semantic and Pragmatic Aspects. Reidel, Dordrecht (1986)
3. Hajičová, E.: Corpus annotation as a test of a linguistic theory: The case of Prague Dependency Treebank, pp. 15–24. Franco Angeli, Milano (2007)
4. Lopatková, M., Plátek, M., Kuboň, V.: Modeling Syntax of Free Word-Order Languages: Dependency Analysis by Reduction. In: Matoušek, V., Mautner, P., Pavelka, T. (eds.) TSD 2005. LNCS (LNAI), vol. 3658, pp. 140–147. Springer, Heidelberg (2005)
5. Lopatková, M., Plátek, M., Sgall, P.: Towards a Formal Model for Functional Generative Description: Analysis by Reduction and Restarting Automata. The Prague Bulletin of Mathematical Linguistics 87, 7–26 (2007)
6. Tesnière, L.: Eléments de syntaxe structurale. Librairie C. Klincksieck, Paris (1959)
7. Mel'čuk, I.A.: Dependency in language. In: Proceedings of DepLing 2011, Barcelona, pp. 1–16 (2011)
8. Gerdes, K., Kahane, S.: Defining dependencies (and constituents). In: Proceedings of DepLing 2011, Barcelona, pp. 17–27 (2011)
9. Jančar, P., Mráz, F., Plátek, M., Vogel, J.: On monotonic automata with a restart operation. Journal of Automata, Languages and Combinatorics 4, 287–311 (1999)
10. Otto, F.: Restarting Automata. In: Reichel, H. (ed.) FCT 1995. LNCS, vol. 965, pp. 269–303. Springer, Heidelberg (1995)

11. Plátek, M., Mráz, F., Lopatková, M.: (In)Dependencies in Functional Generative Description by Restarting Automata. In: Proceedings of NCMA 2010, Wien, Austria, Österreichische Computer Gesellschaft. books@ocg.at, vol. 263, pp. 155–170 (2010)
12. Avgustinova, T., Oliva, K.: On the Nature of the Wackernagel Position in Czech. In: Formale Slavistik, pp. 25–47. Vervuert Verlag, Frankfurt am Main (1997)
13. Hana, J.: Czech Clitics in Higher Order Grammar. PhD thesis, The Ohio State University (2007)
14. Hajičová, E., Havelka, J., Sgall, P., Veselá, K., Zeman, D.: Issues of Projectivity in the Prague Dependency Treebank. The Prague Bulletin of Mathematical Linguistics 81, 5–22 (2004)
15. Holan, T., Kuboň, V., Oliva, K., Plátek, M.: On Complexity of Word Order. Les grammaires de dépendance – Traitement automatique des langues (TAL) 41, 273–300 (2000)
16. Pajas, P., Štěpánek, J.: System for Querying Syntactically Annotated Corpora. In: Proceedings of the ACL-IJCNLP 2009 Software Demonstrations, pp. 33–36. ACL, Singapore (2009)
17. Pajas, P., Štěpánek, J.: Recent Advances in a Feature-Rich Framework for Treebank Annotation. In: Proceedings of CoLING 2008, vol. 2, pp. 673–680. The Coling 2008 Organizing Committee, Manchester (2008)
18. Mikulová, M., Bémová, A., Hajič, J., Hajičová, E., Havelka, J., Kolářová, V., Kučová, L., Lopatková, M., Pajas, P., Panevová, J., Razímová, M., Sgall, P., Štěpánek, J., Urešová, Z., Veselá, K., Žabokrtský, Z.: Annotation on the tectogrammatical level in the Prague Dependency Treebank. Annotation manual. Technical Report 30, Prague, Czech Rep. (2006)