On an Apparent Freedom of Czech Word Order. A Case Study

Kateřina Rysová, Jiří Mírovský and Eva Hajičová

Institute of Formal and Applied Linguistics
Faculty of Mathematics and Physics
Charles University in Prague
E-mail: {rysova|mirovsky|hajicova}@ufal.mff.cuni.cz

Abstract

The aim of the present contribution is to document, on the material of the Prague Dependency Treebank (PDT), that the assumed freedom of Czech word order is not really a freedom but that it is guided by certain principles, different from the grammatically given principles determining the word order in some other European languages such as English, German or French. After a short introductory section summarizing the traditional views on Czech word order (Section 1) we briefly characterize our approach to the information structure of the sentence (TFA) and its representation in the annotated corpus of Czech (PDT, Section 2). In Section 3 we present the results of the automatic procedure for dividing the sentence into its topic and focus part and compare these results with human annotators decisions. In Section 4 we put forward and test the hypothesis on the order of elements in the focus part of the sentence, adding some observations that follow from the tests. A summary of our investigations is given in Section 5.

1 Traditional views on Czech word order

Since the pioneering studies of Vilém Mathesius [17] topic–focus articulation is considered to be the primary factor determining the word order in Czech; other factors influencing the Czech word order are then considered to be secondary, be it the grammatical or the rhythmical factors (see [5, p. 609]). According to Uhlířová [23], among the word order principles governed by grammar there is the basic position of a congruent attribute before the governing noun and the modificiation of manner before the verb; grammatical as well as rhythmical aspects govern the position of the so-called clitics, typically the second sentence position, i.e. the position after the first member of the sentence carrying the stress (the so-called Wackernagel position). A special attention is paid to the position of the so-called rhematizers; their position is closely related to their scope.

2 Topic–focus articulation of the sentence and its representation in PDT

2.1 Topic–focus articulation

Our investigation follows the **theoretical account of topic–focus articulation** (TFA in the sequel, see e.g. [21], [12]) within the framework of the Functional Generative Description, according to which the sentence can be divided into what the sentence is **about** (its topic) and what it says **about the topic** (its focus). It is assumed that the **dichotomy of topic and focus** (which is supposed to be very important especially for the specification of the scope of negation) is based on the primary notion of contextual boundness.

2.2 TFA in the Prague Dependency Treebank

The **empirical material** we base our analysis on is the (mostly) manually annotated corpus of Czech, the **Prague Dependency Treebank** (PDT, [2]). TFA is captured there by means of a special attribute of TFA assigned to (almost) each node of the deep structure dependency tree (so-called tectogrammatical level) which may obtain one of the three values: *t* for a non-contrastive contextually bound node, *c* for a contrastive contextually bound node and *f* for a contextually non-bound node.[1] It is assumed that the verb stands on the boundary between topic and focus. The left-to-right dimension of a tectogrammatical tree serves as the basis for the specification of the scale of communicative dynamism: communicative dynamism is understood as the deep word order, with the dynamically lowest element standing in the leftmost position and the most dynamic element (the focus proper of the sentence) as the rightmost element of the dependency tree.

3 Automatic procedure of topic–focus division

3.1 Algorithm for topic–focus division

The algorithm determining the topic–focus boundary is based on the information on the contextual boundness for each node of the tectogrammatical tree and takes into account the status of the main verb (i.e. the root) of the sentence and its immediate dependents.[2] Basically, the algorithm (referred to in the sequel as SH algorithm) includes the following steps ([22], [20]):

[1] There are 206,537 tectogrammatical nodes annotated as contextually bound, out of them 30,312 are contrastive contextually bound. Further, 354,841 nodes are contextually non-bound and for 41,332 nodes, contextual boundness is not annotated (e.g., coordinating nodes, nodes inserted for technical reasons etc.).

[2] Another algorithm (see [14]) for the identification of the boundary between topic and focus in Czech, different from the one discussed below, was based on the position of the verb and the order of verb complementations following the verb compared to the hierarchy of systemic ordering (this hierarchy is discussed below in Section 4).

(a) the main verb (V) and any of its direct dependents belong to focus iff they have the TFA value index *f*;

(b) every item that does not depend directly on V and is subordinated to an element of focus as determined by (a), belongs to focus (where "subordinated to" is defined as the irreflexive transitive closure of "depend on");

(c) iff V and all items directly depending on V are *t* or *c*, then follow the rightmost edge leading from the main verb to the first node(s) on this path that is/are contextually non-bound; this/these node(s) and all the nodes subordinated to it/them belong(s) to focus;

(d) every item not belonging to focus according to (a) – (c) belongs to topic.

3.2 The original implementation

Zikánová et al. [25] implemented and applied the SH algorithm to a part of the PDT data (about 11 thousand sentences). The results of their implementation indicate that a clear division of the sentence into topic and focus according to the hypothesized rules has been achieved in 94.28% of sentences to which the procedure has been applied; 4.41% of sentences contained the type of focus referring to a node (or nodes) that belong(s) to the communicatively most dynamic part of the sentence though they depend on a contextually bound node. The real problem of the algorithm then rests with the case of ambiguous partition (1.14%) and cases where no focus was recognized (0.11%).

To validate the algorithm, the same PDT documents were analyzed manually, most of them in three parallel annotations (about 10 thousand sentences), and about 600 sentences in six parallel annotations [26]. The annotators (mostly high school students, having some basic idea of the dichotomy of topic and focus as "the aboutness relation" but not being familiar with the theoretical framework of TFA) worked with the raw texts and were instructed to mark – according to their understanding of the given sentence – every single word in the sentence as belonging either to topic or to focus, and, in addition, to mark which continuous part of the sentence they understand as topic and which continuous part as focus. One of the important subtasks of this project was to follow annotators' agreement/disagreement. The disagreement in the assignments of the two parts of the sentence as a whole was rather high and indicates that the intuitions concerning the division of the sentence into its topic and focus parts may dramatically differ. It is interesting to note that the annotators' agreement in the assignments of individual words in the sentences to topic or to focus was much higher (75%) than the agreement in the assignment of the topic-focus boundary (36%) in both the three and six parallel analyses. Most of the cases in which the annotators disagree concerned the position of the verb in the topic or in the focus. It should be also taken into consideration that while we get only a single, unambiguous result from the automatic procedure, more ways of interpretation could be correct.

3.3 New experiment setting and implementation of the algorithm

In the present study, we evaluate the SH algorithm for the division of the sentence into its topic and focus part presented above in Section 3.1 in a slightly different way. First of all, we tried to avoid some of the shortcomings of the implementation of the algorithm reported in [25], most importantly the inability to properly treat coordination structures. Another difference was that we used data annotated by a linguistic expert as the gold data rather than those used in [23] as a result of voting based on the agreement/disagreement of annotators. Our assumption was that both these changes in the experiment setting would bring an improvement in the results and would better reflect the adequacy of the algorithm for transforming values of contextual boundness into the division of the sentence into the topic and the focus.[3]

Our gold data consisted of 319 sentences from twelve PDT documents annotated by a single linguistic expert familiar with the topic–focus articulation theory. The annotation proceeded directly on the tectogrammatical trees, in an annotation environment adjusted for the task. Without taking into account (already annotated but now hidden) values of contextual boundness, the annotator marked each tectogrammatical node as belonging either to the topic or to the focus.

On these gold data, we evaluated our implementation[4] of the SH algorithm, as well as a baseline algorithm similar to the baseline used in [25].[5] However, in our case both the implementation of the SH algorithm and the implementation of the baseline algorithm took into account coordination structures and the rules were applied to each coordinated member separately. Similarly to [25], we did not count all surface sentence tokens in the evaluation, but rather – in our setting – only nodes relevant for topic–focus articulation assignment.[6]

Table 1 shows a comparison of our implementation of the baseline algorithm and our implementation of the SH algorithm. Our implementation of the SH algorithm significantly outperforms the baseline in most of the measured phenomena.

At the same time, both our implementations of the baseline algorithm and of the SH algorithm outperform most of the results presented in [25]; their F1-measure in focus was 0.88 for the baseline and 0.83 for their implementation of the SH algorithm (they did not report their results in topic). The results are strictly speaking not directly comparable, as the implementations were evaluated on different data and in slightly different experiment settings. However, if we exclude the proper treatment of coordinations in our implementation (thus getting closer to the implementation used in [25]), our results for F1-measure in focus drop to 0.8 for the

[3] By using the expertly annotated data we have lost the connection with the language intuition of the non-expert annotators. In this sense, we evaluate the algorithm rather than the agreement between the TFA theory and the intuition of non-expert speakers; that is exactly what we wanted to do.

[4] We use a slightly modified implementation of the SH algorithm programmed in 2007 by Jiří Havelka, which – to our best knowledge – has never been published.

[5] The baseline is defined as follows: in the linear (surface) form of the sentence, each word before the autosemantic part of the predicate verb belongs to topic, the rest of the sentence belongs to focus.

[6] I.e. we only evaluated tectogrammatical nodes that had a value of contextual boundness filled in. It means that for example coordination nodes have been left out of the evaluation.

Measure	Baseline	SH Algorithm
recall in topic	0.69	0.94
precision in topic	0.89	0.85
F1-measure in topic	0.78	0.89
recall in focus	0.96	0.93
precision in focus	0.88	0.97
F1-measure in focus	0.92	0.95
overall accuracy on tectogrammatical nodes	0.88	0.93
overall accuracy on whole sentences	0.31	0.75

Table 1: Evaluation of our implementation of the baseline algorithm and of our implementation of the SH algorithm on our gold data.

baseline algorithm and to 0.88 for the SH algorithm. It means that the proper way of processing coordinations plays an important factor in achieving better results. Apart from this, we attribute the improvement also to the different kind (we believe "better quality") of the gold data on which we measure the results. We find also very important that our implementation of the SH algorithm outperforms the baseline, while in the experiments reported in [25], the results were the opposite.

4 Order of words in the focus part of the sentence

4.1 General remarks

As stated in Section 1, the topic–focus articulation is traditionally understood as the primary factor of Czech word order. However, with a more detailed look at the word order in relation to the topic–focus bi-partition, a substantial difference has come out. The ordering of the elements in the topic part is basically motivated by the degree of activation of the corresponding elements of the stock of knowledge assumed by the speaker to be shared by him and the hearer,[7] while in the focus part there is a close relationship of the ordering of the sentence elements to the syntactico-semantic types of the verb modifications.

The postulation of a certain hierarchy of cognitive entities that is reflected in the ability of the hearer to identify the referents of the expressions used in the sentence and eventually reflected in the order of sentence elements dates back to the eighties of the last century. Based on the notion of "assumed familiarity", a familiarity scale is defined by Prince [18, p. 245]. In a similar vein, Gundel et al. [11] accepts Givon's [6] scale in the syntactic coding of topic accessibility and maintains that the assumed cognitive (memory and attention) status of an intended referent/interpretation is connected with the appropriate use of a different form or

[7] For the notion of this hierarchy see e.g. [15], [13].

forms (articles, demonstrative pronouns). Ariel [1] concentrates on the system of accessing NP antecedents and follows Givon's view that the Accessibility Marking Scale cannot be taken as universal because it does not cover the full range of referring expressions in all languages. Lambrecht [16] bases his analysis of referent accessibility on Chafe's ([3], [4]) idea of three activation states and relates them to their formal correlates in the structure of sentences. A slightly different but yet related is the model of the local attentional states of speakers and hearers as proposed by Grosz and Sidner ([9], [10]), which is the basis of the centering theory ([7], [8], [24]). Discourse entities are ranked according to language-specific ranking principles; the ranking of centers is defined in terms of syntactic relations as specified for the surface shape of the sentence (subject, object etc.) and as such are very language-dependent.

Considerations similar to these have led us to study the word order separately for the topic and for the focus part of the sentence. Apart from a minor influence of grammatical principles mentioned in Section 1, we consider the order of sentence elements in the **topic** part to be basically influenced by the previous co-text and the situational contexts as well as by the intended discourse strategy of the speaker, his interests and also by his intention to put certain elements in contrast to the previous co-text or situation. For the order of elements in the focus part of the sentence, a hypothesis of the so-called systemic ordering was formulated; it is discussed in more detail below.

4.2 The hypothesis of systemic ordering

The original empirical study of Czech texts has led to the assumption [21, p. 69] that the ordering of the elements in the **focus** part of the sentence is primarily given by the type of the complementation of the verb. This hypothesis was empirically tested pairwise (i.e., successively for two of the complementation types) and it was also supported by several psycholinguistic experiments [21, p. 72ff]. It was assumed that systemic ordering is a universal phenomenon and that at least in most European languages the order of the principle verb complementations (such as Actor – Addressee – Patient) is the same. The following detailed ordering has been established for Czech:

Actor – Temporal (when – since when – till when – how long) – Location (where) – Manner – Extent – Measure – Means – Addressee – From where – Patient – To where – Effect – Condition – Aim – Cause

4.3 Testing of the hypothesis

It was clear from the very beginning that the hypothesis of systemic ordering is very **strong** and that further investigation based on a much broader material is needed, which may lead to a more precise specification or modification(s). The material of the **Prague Dependency Treebank** opened the possibility to validate

the hypothesis. A rather complex analysis is presented by K. Rysová [19], who arrives (among other important findings) at the following observations:

(i) There is a tendency of a contextually non-bound element expressed by a clause to follow the non-sentential element (which is apparently connected with the 'weight' of the element).

(ii) There is an influence of the form of the complementation: e.g., the assumed order Manner – Patient is more frequent if the complementation of Manner is expressed by an adverb and the complementation of Patient by a nominal group.

(iii) As for the position of the Actor on the scale, a substantial number of counterexamples of the original hypothesis concern cases for which the outer form of the Actor plays an important role: in sentences with the verb *být* [*to be*] in structures of the type *je nutné* (PAT) *přiznat* (ACT) [*it is necessary* (PAT) *to acknowledge* (ACT)], where Actor is expressed by infinitive, Patient precedes Actor, while the hypothesized order Actor – Patient is attested to if both complementations are expressed by nominal groups.

(iv) There is also a possibility that the order might be influenced by the difference in the optional/obligatory character of the given complementations: there is a tendency that obligatory complementations follow the optional ones though this tendency is not a very influential word order factor.

Rysová's analysis confirms that there is a considerable tendency that in such pairs one ordering prevails over the other, which was the starting point of the postulation of the systemic ordering hypothesis. However, with some pairs, such as Patient and Means, there was a balance between the frequency of the two possible orders, which may indicate that for some particular complementations more than a single complementation occupy one position on the scale.

4.4 Further supporting factors

4.4.1 Position of the verb

The data on the annotators' agreement/disagreement on the topic/focus boundary presented above in Section 3.2 have invited a more detailed study of the decisions of the annotators, especially from the point of view of a possible influence of the verb position.

It has been confirmed that in the Czech surface word order the verb (be it contextually bound or non-bound) can be shifted into the **second position** even if followed by contextually bound elements: both Examples (1) and (2) can serve as a reply to *What did Dan do yesterday?* or *Where did Dan go yesterday?*, i.e the verb *jel* [*went*] may be contexually bound or non-bound.

(1) *Dan včera jel z Prahy do Brna.*
[Lit. *Dan yesterday went from Prague to Brno.*]

(2) *Dan jel včera z Prahy do Brna.*
[Lit. *Dan went yesterday from Prague to Brno.*]

In the sample of the PDT studied for this purpose,[8] there are 6,458 sentences (15% cases) with the verb in the second position as compared with 37,208 sentences with the verb on other than the second position. The order of the contextually non-bound complementations of the verb was studied pairwise, with the following results:[9]

(a) The ordering was in accordance with the hypothesized systemic ordering in cases of LOC–PAT, PAT–AIM, TWHEN–ADDR, LOC–EXT, PAT–EFF, TWHEN–PAT, ADDR–PAT, THO–PAT, THL–PAT and TWHEN–LOC; this result holds for the sentences regardless of the verb position in them, see Example (3).

(3) *Termín pseudohumanisté má ostatně v našem politickém životě*.LOC *dlouhou tradici*.PAT
[Lit. *The term pseudohumanists has, after all, in our political life*.LOC *a long tradition*.PAT]

(b) The hypothesized systemic ordering was confirmed also with the pairs EXT–PAT, TWHEN–EXT and MANN–EFF. However, in sentences with the verb on other than the 2nd position, the supposed ordering predominated very strongly, except for the pair EXT–PAT which was present especially with the verb on the 2nd position. For pairs MANN–PAT (see Example (4)) and PAT–DIR3, the systemic ordering was confirmed in sentences with the verb on other than the 2nd position. In cases with the verb on the 2nd position, none of the orderings was prevailing.

(4) *Kromě toho do tří let postaví ve městě na své náklady*.MANN *travnaté fotbalové hřiště*.PAT
[Lit. *In addition, within three years, (they) will build in the city at their expense*.MANN *a grass football field*.PAT]

(c) There was a considerable difference between the cases with the verb on the 2nd position and cases with the verb in other than the 2nd position in the pair PAT–DIR1. In sentences with the verb on other than the 2nd position, the hypothesized ordering DIR1–PAT was present more frequently (see Example (5)), while in sentences with the verb on the 2nd position, the ordering PAT–DIR1 strongly predominated (see Example (6)).

[8] The sentences studied included the verb in the second position in the surface shape of the sentence when the verb was followed by a dependent that, in the corresponding tectogrammatical representation, had the *c* or *t* value of the TFA attribute. Only sentences with indicative mood were taken into account. For testing our hypotheses, 9/10 of the PDT data have been used.

[9] In order to exclude the factor of "weight", we have taken into account only clauses in which the predicate had any number of dependents but the relevant dependents had maximally 3 subordinated nodes.

(5) *Ve starých filmech s Oldřichem Novým hlavy majetných rodin snímaly ze stěn*.DIR1 *obrazy*.PAT*, aby ze schránek za nimi vylovily notářské listiny či rodinné šperky.*
[Lit. *In old movies with Oldřich Nový, the heads of the wealthy families removed from the walls*.DIR1 *paintings*.PAT *in order to find some notarial deeds or the family jewels in the boxes behind them.*]

(6) *Stuttgartská automobilka Porsche přesune letos výrobu*.PAT *ze svého závodu*.DIR1 *v Salcburku do Českého Krumlova.*
[Lit. *The Stuttgart car maker Porsche will move this year the production*.PAT *from its factory*.DIR1 *in Salzburg to Český Krumlov.*]

This analysis has pointed out that in most cases the 2nd position of the verb does not influence the order of the contextually non-bound complementations of the verb. However, the results of this follow-up analysis have also confirmed that the position of ACT in the systemic ordering hierarchy (first on the scale, before all other types of complementations) has to be revised, or at least the contextual conditions for its appearance in other positions in the focus part of the sentence have to be studied in more detail. The pairs that documented this necessity are PAT–ACT, TWHEN–ACT, EXT–ACT, MANN–ACT and THO–ACT. Also the ordering in pair MANN–LOC has to be further analyzed.

4.4.2 Word order in separate clauses

In the original formulation of the systemic ordering as well as in its testing on the PDT no difference was made between the ordering in the **main clause** and in the **dependent clauses**.

We have therefore performed a broader research on testing the hypothesis of systemic ordering with respect to this factor and studied the word order of contextually non-bound verb complementations in the main clauses and in the dependent clauses separately.

This analysis has pointed out that in most cases the sentence character (main clause vs. dependent clause) does not influence the order of the contextually non-bound complementations of the verb within the respective clauses. The main results of the analysis are as follows:

(a) The ordering was in accordance with the hypothesized systemic ordering in case of TWHEN–ADDR, ACT–ADDR, PAT–AIM, MANN–EFF and DIR1–PAT; this result holds both for main clauses and for dependent clauses.

(b) The hypothesized systemic ordering was confirmed also in pairs PAT–EFF, EXT–PAT and TWHEN–EXT in both types of clauses. Pairs PAT–EFF and EXT–PAT occurred in dependent clauses significantly more often, while TWHEN–EXT occurred rather in main clauses. The supposed ordering PAT–DIR3 and LOC–EXT was confirmed only in main clauses. In dependent clauses, the ordering in these pairs was nearly balanced.

As in Section 4.4.1, the analysis has also indicated that the position of ACT in the systemic ordering hierarchy has to be studied in more detail, considering the frequently occurring pairs PAT–ACT and MANN–ACT. Also the ordering in pairs PAT–MEANS and MANN–LOC has to be further analyzed.

We have also followed the appurtenance of the dependent clause to the topic or to the focus part of the sentence, i.e. whether the main predicate of the given dependent clause was contextually bound or non-bound. Unfortunately, the number of contextually bound dependent clauses was very small when compared to the number of dependent clauses that are part of focus; out of 2,861 dependent clauses (fulfilling the above restrictions) there were only 141 dependent clauses the predicate of which had the value of the TFA attribute *t* or *c* (i.e. contextually bound). This low occurrence of dependent clauses in the topic part of the sentence does not allow to draw any conclusions on the order of the verb complementations in them, except for the statement that our data indicate that most dependent clauses occur in the focus part of the sentence.

5 Summary

Testing the relation between TFA and Czech word order on the annotated corpus of Czech, the following assumptions have been confirmed and the following tendencies have been observed:

(a) Czech word order is not ultimately free but it is guided by **communicatively** given principles.

(b) The order of elements in the focus part of the sentence reflects in principle the hypothesis of **systemic ordering**.

(c) There are several factors influencing systemic ordering following from the **syntactic** structure of the sentence and the **lexical** properties of sentence elements.

(d) In most cases, neither the sentence character (the **main** clause **vs. dependent** clause) nor the position of the governing verb in the sentence (the position on the **2nd place vs. other positions**) influence the order of the contextually non-bound complementations of the verb within the sentence.

Our **implementation of the SH algorithm** for the division of the sentence into topic and focus based on the values of contextual boundness outperforms previously published results and, also very importantly, outperforms the baseline.

Acknowledgements

We gratefully acknowledge support from the Ministry of Education of the Czech Republic (project LM2010013 – LINDAT/CLARIN) and the Czech Science Foundation (GA CR, project P406/12/0658). This work has been using language resources developed, stored and distributed by the LINDAT/CLARIN project.

References

[1] Ariel M. (1990). *Accessing Noun-Phrase Antecedeents.* London: Routledge.

[2] Bejček E., E. Hajičová, J. Hajič, P. Jínová, V. Kettnerová, V. Kolářová, M. Mikulová, J. Mírovský, A. Nedoluzhko, J. Panevová, L. Poláková, M. Ševčíková, J. Štěpánek and Š. Zikánová (2013). *Prague Dependency Treebank 3.0.* Data/software, Univerzita Karlova v Praze, MFF, ÚFAL, Prague.

[3] Chafe W. (1979). The flow of thought and the flow of language. In: Givon (ed.), *Syntax and Semantics: Discourse and Syntax, Vol. 12.* New York: Academic Press, 159–182.

[4] Chafe, W. L. (1987). Cognitive constraints on information flow. In: R. S. Tomlin (ed.), *Coherence and grounding in discourse.* Amsterdam: Benjamins, 21–52.

[5] Daneš F., H. Běličová, M. Čejka, E. Dvořák, M. Grepl, K. Hausenblas, Z. Hlavsa, J. Hoffmannová, J. Hrbáček, J. Chloupek, P. Karlík, E. Macháčková, O. Müllerová, B. Palek, J. Nekvapil, J. Novotný, P. Piťha, H. Prouzová, M. Rulfová, B. Rulíková, O. Šoltys, L. Uhlířová and S. Žaža (1987). *Mluvnice češtiny. 3. Skladba* [*Grammar of Czech. 3. Syntax*]. Prague: Academia.

[6] Givon T. (1983). Topic continuity in discourse: An Introduction. In: Givon (ed.), *Topic Continuity in Discourse: A Quantitative Cross-Language Study.* Amsterodam: John Benjamins, 1–41.

[7] Grosz B., A. K. Joshi and S. Weinstein (1983). Providing a unified account of definite noun phrases in discourse. In: *Proceedings of the Annual Meeting of the Association for Computational linguistics 21,* 44–50.

[8] Grosz B., A. K. Joshi and S. Weinstein (1995). Centering: A Framework for modeling the local coherence of discourse. *Computational Linguistics 21,* 203–225.

[9] Grosz B. J. and C. L. Sidner (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics 12,* 175–204.

[10] Grosz B. J. and C. L. Sidner (1998). Lost intuitions and forgotten intentions. In: M. Walker, A. Joshi and E. Prince (eds.), *Centering in Discourse.* Oxford University Press, 39–51.

[11] Gundel J. K., N. Hedberg and R. Zacharski (1993). Cognitive status and the form of referring expressions in discourse. *Language 69,* 274–307.

[12] Hajičová E., B. Partee and P. Sgall (1998). *Topic–Focus Articulation, Tripartite Structures and Semantic Content.* Dordrecht: Kluwer.

[13] Hajičová E. (1993). *Issues of Sentence Structure and Discourse Patterns.* Prague: Charles University Press.

[14] Hajičová E., H. Skoumalová and P. Sgall (1995). An Automatic Procedure for Topic–Focus Identification. *Computational Linguistics, 21(1)*, 81–94.

[15] Hajičová E. and J. Vrbová (1982). On the role of the hierarchy of activation in the process of natural language understanding. In: Horecký J. (ed.), *Coling 82 – Proceedings of the Ninth International Congress of Computational Linguistics,* Prague – Amsterdam, 107–113.

[16] Lambrecht K. (1994). *Information Structure and Sentence Form. Topic, Focus and the Mental Representations of Discourse Referents.* Cambdridge: Cambridge University Press.

[17] MathesIus V. (1941). Základní funkce pořádku slov v češtině. [Basic functions of word order in Czech.] *Slovo a slovesnost 7,* 169–180.

[18] Prince E. (1981). Toward a taxonomy of given/new information. In: Cole P. (ed.), *Radical Pragmatics.*. New York, Academic Press, 223–254.

[19] Rysová K. (2014). *O slovosledu z komunikačního pohledu.* [*On Word Order from the Communicative Point of View.*] Prague: Institute of Formal and Applied Linguistics.

[20] Stall P. (1981). Towards a Definition of Focus and Topic. In: *Prague Studies in Mathematical Linguistics 7.* Prague: Academia, 173–197.

[21] Sgall P., E. Hajičová and E. Buráňová (1980). *Aktuální členění věty v češtině.* [*Topic–Focus Articulation in Czech.*] Prague: Academia.

[22] Sgall P. and E. Hajičová (1977). Focus on focus. In: *The Prague Bulletin of Mathematical Linguistics, No. 28,* 5–54, and *No. 29* (1978), 23–41. Prague: Univerzita Karlova v Praze.

[23] Uhlířová L. (1987). *Knížka o slovosledu.* [*A book on word order.*] Prague: Academia.

[24] Walker M. A., A. Joshi and E. Prince (1998). Centering in naturally occurring discourse: An overview. In: Walker M. A., A. Joshi and E. Prince (eds.), *Centering theory in discourse.* Oxford: Clarendon, 1–28.

[25] Zikánová Š. and M. Týnovský (2009). Identification of Topic and Focus in Czech: Comparative Evaluation on Prague Dependency Treebank. In: Zybatow, Gerhild et al. (eds.), *Studies in Formal Slavic Phonology, Morphology, Syntax, Semantics and Information Structure.* Formal Description of Slavic Languages 7, Frankfurt am Main: Peter Lang, 343–353.

[26] Zikánová Š., M. Týnovský and J. Havelka (2007). Identification of Topic and Focus in Czech: Evaluation of Manual Parallel Annotations. In: *The Prague Bulletin of Mathematical Linguistics, No. 87.* Prague: Univerzita Karlova v Praze, 61–70.