What do the data in Prague Dependency Treebank say about systemic ordering in Czech?

Šárka Zikánová

1. Introduction

Since ancient linguistics, the studies of Indo-European word order work with the conception of universal natural word order (ordo naturalis) – an order of the verb-dependent constituents in the linear organization of a clause. The description of the natural word order is usually based on occasional (and in some degree random) observations of clauses in a certain language.

In Czech linguistics, the idea of the natural word order was formulated in a more precise way as the hypothesis of the "systemic ordering" (Sgall, Hajičová and Buráňová, 1980). According to the authors, the contextually non-bound participants and adverbials are ordered as follows (o. c., page 77):

actor – time (when – since when – to when – how often – how long) – location (where) – manner – measure – instrument (mean) – direction (which way) – addressee – origine – direction (from where) – patient – direction (where) – effect – condition – aim – reason

(This formulation of the hypothesis in the 1970s was based on an analysis of hundreds of sentences and was supported by psycholinguistic experiments.)

In this paper, we check a part of the hypothesis, namely the placement of the actor, addressee and patient in Czech, on data from the Prague Dependency Treebank. The Prague Dependency Treebank contains 800 000 words with a complex syntactic and morphemic annotation together with an annotation of the information structure. This allows to classify the deviations from the systemic ordering as exhibited in the real data and especially to describe regularities in the deviations.
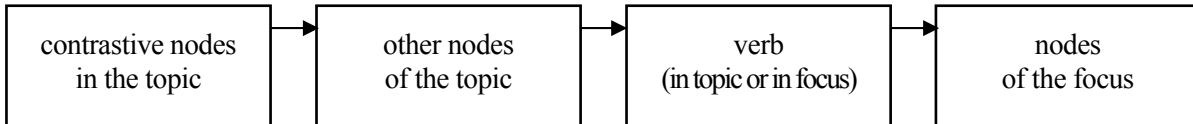
2. The hypothesis of the systemic ordering in Czech

The hypothesis of the systemic ordering (SO) in Czech is a part of the theory of the topic-focus articulation (cf. Sgall, Hajičová and Buráňová 1980) within the general theoretical frame of the Functional generative description of language (Sgall 1967, Panevová 1980, Sgall, Hajičová and Panevová 1986).

In the Functional generative description, the structures of clauses are described on the layer of deep (underlying) syntax in syntactic dependency trees (tectogrammatical level). The trees consist of nodes in which each node represents a semantic unit corresponding approximately to one lexically autonomous (autosemantic) word on the surface layer. Each node has its specific role in the topic-focus articulation of a sentence. The primary notion on which the bipartition of the sentence into its topic (what the sentence is about) and its focus (what is said about the topic) is based is the notion of contextual boundness: contextually bound elements occur primarily (but not only) in the topic, contextually non-bound elements occur prototypically in the focus (but not only there; the secondary cases concern deeper embedded tokens). In the PDT scenario, there is a specific TFA (topic-focus articulation) attribute with three possible values (Mikulová et al. 2005, page 1120):

t – a non-contrastive contextually bound node
c – a contrastive contextually bound node
f – a contextually non-bound node

The nodes are ordered in a syntactic tree from the left to the right according to the rising communicative dynamism. Prototypically, the contextually bound nodes precede the contextually non-bound nodes; the border node between the contextually bound and contextually non-bound parts of a sentence is the node of the predicate verb. Thus, the so called deep word order looks as follows:

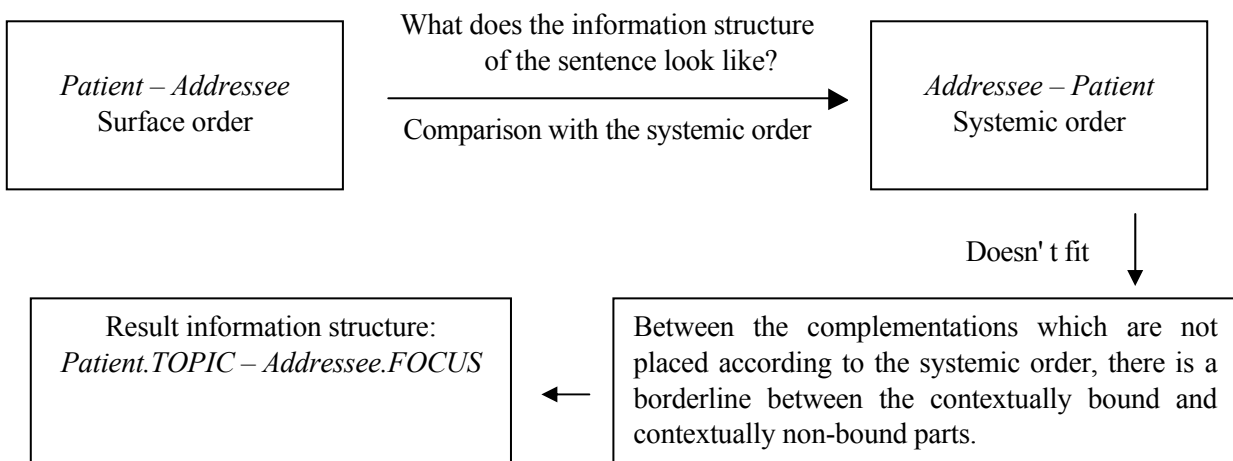| contrastive nodes in the topic	→	other nodes of the topic	→	verb (in topic or in focus)	→	nodes of the focus

The hypothesis of the systemic order concerns the underlying placement of the nodes in focus immediately dependent on the verb. According to the hypothesis, the verbal complementations in focus have a strict order which follows the rising communicative dynamism, e.g. addressee – patient (Sgall, Hajičová and Buráňová 1980, page 70).

> *Dával děvčatům.*ADDR $_{FOCUS}$ *květiny.*PAT $_{FOCUS}$
> *He-gave girls.*DAT. PL. *flowers.*ACC. PL.

Other surface orders of the verbal complementations are not excluded, but they are not synonymous. Often they correspond to different scales of communicative dynamism and they indicate that the sentence has another information structure (and subsequently another meaning).

> *Dával květiny.*PAT $_{TOPIC}$ *děvčatům.*ADDR $_{FOCUS}$
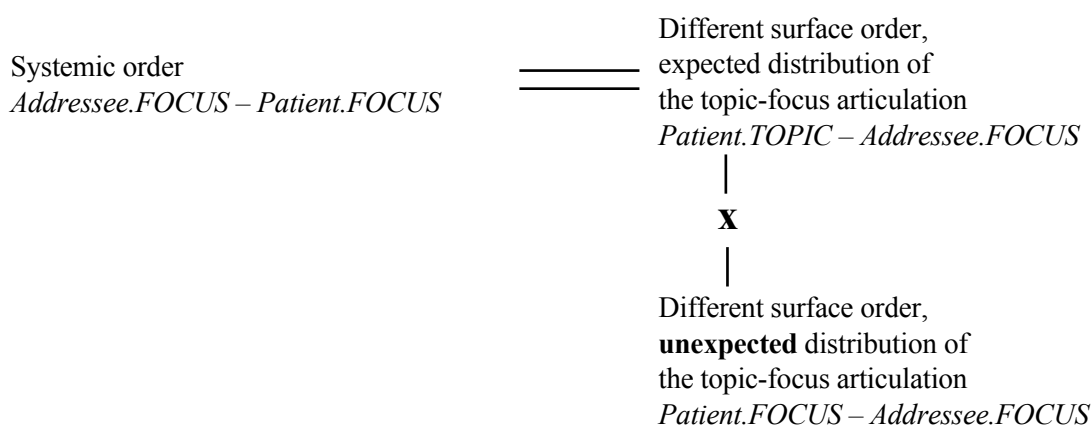> *He-gave (the) flowers to (–) girls.*

It follows from the hypothesis described above that in sentences with a surface order different from the systemic ordering, it is possible to identify the borderline between the contextually bound and contextually non-bound parts of a sentence. Whenever a verbal complementation on the right side of a sentence[1] precedes another one in contradiction with the systemic order, it is contextually bound whereas the following verbal complementation is contextually non-bound.

Patient – Addressee Surface order	What does the information structure of the sentence look like? ——————————— Comparison with the systemic order →	*Addressee – Patient* Systemic order

Doesn' t fit ↓

Result information structure: *Patient.TOPIC – Addressee.FOCUS*	←	Between the complementations which are not placed according to the systemic order, there is a borderline between the contextually bound and contextually non-bound parts.

---

[1] The statement about the borderline between the contextually bound and contextually non-bound parts of a sentence applies to the rightmost violation of the SO in the sentence (cf. the following illustration). Let us have a SO of five complementations: 1 2 3 4 5 and the following surface order of the five complementations: 1 3 2 5 4. According to the statement, there is a borderline between the complementations 5 and 4 (5 is contextually bound, 4 is contextually non-bound). But, we cannot in this way set the value of the topic-focus articulation in the pair of complementations 3 and 2 which do not correspond with the SO as well. Either the complementation 3 is contextually bound and 2 is contextually non-bound, or both of them are contextually bound.

3. Testing the SO hypothesis on the Prague Dependency Treebank

For the purposes of our analysis, we have excerpted sentences from the Prague Dependency Treebank contradicting the SO hypothesis, i.e. sentences in which the surface word order doesn't correspond to the systemic order. In such cases, according to the hypothesis, the first verbal complementation in a critical pair of complementations should be contextually bound. Nevertheless, in some examples this verbal complementation has been annotated as a contextually non-bound one.

Systemic order
Addressee.FOCUS – Patient.FOCUS

———————

Different surface order,
expected distribution of
the topic-focus articulation
Patient.TOPIC – Addressee.FOCUS

X

Different surface order,
unexpected distribution of
the topic-focus articulation
Patient.FOCUS – Addressee.FOCUS

The analysis deals with the positions of actor, addressee and patient in Czech. The sentences we are concerned with, exemplify the following possible cases:

Patient.FOCUS – Addressee.FOCUS
Addressee.FOCUS – Actor. FOCUS
Patient.FOCUS – Actor. FOCUS

4. The main types of sentences contradicting the hypothesis of the systemic ordering

Disregarding the group of sentences with an erroneous annotation of the topic-focus articulation, six main groups of sentences violating the systemic order arise in our material. In some of them, the deviation seems to be caused by further factors of the topic-focus articulation (4.1., 4.2.); in others, the violation is connected with the form of the verbal complement (4.3., 4.4., 4.5.) or it has theory internal reasons (4.6.).

4.1. The right-hand argument is a focus proper, the left one is still contextually non-bound

As the examples (1–3) demonstrate, in some constructions verbal complementations in focus are inverted contrary to the systemic order, but still according to the rising communicative dynamism. The right-hand member of the pair constitutes the focus proper (a piece of the most important new information in a sentence, a node with the highest level of the communicative dynamism; cf. Sgall, Hajičová and Buráňová 1980, page 51), therefore it is placed in the rightmost surface position in the sentence. Nevertheless, the left member of the pair is not contextually bound. In further work, these sentences must be described and explained in a more detailed way and must be taken into account as an exception from the hypothesis of systemic order.

(1) PAT – ACT

V jisté ulici na pražských Královských Vinohradech si před několika měsíci otevřel soukromou ordinaci.PAT FOCUS psychiatr.ACT FOCUS

  lit.: *In certain street in Praguian Královské Vinohrady some months ago set up private practice.PAT FOCUS, ACC. SG. psychiatrist.ACT. FOCUS, NOM. SG.*

  Some months ago a psychiatrist set up a private practice in a certain street of Praguian Královské Vinohrady.

(2) ADDR – ACT

Rusko, Velká Británie a Francie údajně vypracovaly nový mírový plán pro bývalou Jugoslávii, s nímž uplynulý víkend seznámil srbského prezidenta Slobodana Miloševiče.ADDR FOCUS ruský ministr zahraničí Andrej Kozyrev.ACT FOCUS.

  lit.: *Russia, Great Britain and France reportedly prepared new peace plan for former Yugoslavia, about which last weekend informed Serbian president Slobodan Milošević.ADDR FOCUS, ACC. SG. Russian foreign minister Andrey Kozyrev.ACT FOCUS, NOM. SG.*

  Russia, Great Britain and France are reported to have prepared a new peace plan for the former Yugoslavia, which was presented to the Serbian president Slobodan Milošević by Russian foreign minister Andrey Kozyrev last weekend.

(3) PAT – ADDR

*Průtrž mračen, která postihla Krkonoše z pátku na sobotu, způsobila **značné škody.PAT** Krkonošskému národnímu parku.ADDR.*

  lit.: *Cloud-burst, which afflicted Krkonoše from Friday to Saturday, caused considerable damage.PAT FOCUS, ACC. SG. to_Krkonoše National Park.ADDR FOCUS, DAT. SG.*

  A cloud-burst which afflicted the Krkonoše mountains on Friday night caused considerable damages to the Krkonoše National Park.

4.2. Influence of a contrastive bound complement at the beginning of a sentence

In the following group of examples, the violation of the systemic order seems to be connected with the presence of a contrastive element at the beginning of the sentence. The initial contrastive element has a strong accent[2] which probably influences the intonational curve of the sentence and, as we can see in our material, typically brings an actor complementation to the end as an accented counterpart of the contrast at the beginning.

  In Czech, this surface word order is often used in journalistic texts. Such a word order is not neutral, it is expressive and the expressivity is achieved precisely by the violation of the systemic order. Hence these constructions don't contradict the hypothesis at the same level as the examples in (1–3); in this case, the deviation from the norm (systemic order) is used functionally, as a means of expressivity.

(4) ADDR – ACT

Na 300 tisíc marek.CONTRAST ročně přinese Českomoravskému fotbalovému svazu.ADDR FOCUS smlouva.ACT FOCUS s firmou Puma.

  lit.: *About 300 thousand marks.CONTRAST annually will_bring to_Czechomoravian football union.ADDR FOCUS, DAT. SG. contract.ACT FOCUS, NOM. SG. with firm Puma.*

  The contract with the firm Puma will bring about 300 thousand DM annually to the Czechomoravian football union.

---

[2] There is a possibility of interpreting the first complementation in the given examples as a focus proper as well which has been misinterpreted by annotators. In such a case, the pitch accent at the beginning would be especially well explainable. Anyhow, the inversion of the focus complementations on the right side of the verb still remains and should be explained or at least described. (In spoken texts, the right interpretation of the first complementation – contrast vs. focus proper – would be indicated by the intonation.)

(5) PAT – ACT

*Do ponožky.*CONTRAST *si schoval tři a půl tisíce.*PAT FOCUS *korun čtyřicetiletý muž.*ACT FOCUS *z Ostravy.*

lit.: *In sock.*CONTRAST *hid three and half thousand crowns.*PAT FOCUS, ACC. PL. *fourty-year-old man.*ACT FOCUS, NOM. SG.*from Ostrava.*

A fourty-year-old man from Ostrava hid three thousand five hundred crowns in his sock.

(6) PAT – ACT

*Celkem ve 12 sladovnách na Moravě i v Čechách.*CONTRAST *vyrábí slad.*PAT FOCUS *akciová společnost.*ACT FOCUS *Obchodní sladovny Prostějov.*

lit.: *Altogether in 12 malthouses in Moravia and Bohemia.*CONTRAST *produces malt.*PAT FOCUS, ACC. SG.*joint-stock company.*ACT FOCUS, NOM. PL. *Prostějov Trade Malthouses.*

The Prostějov Trade Malthouses, the joint-stock company, produces malt altogether in 12 malthouses in Moravia and Bohemia.

4.3. Phraseme [verb + complement] + other complements

The inversion of the verbal complements in focus is characteristic for sentences with verbs with a general meaning (e.g. *to be, to give, to have*) where the verb and the first complement on the right-hand build a complete semantic unit. Here, the connection between the verb and the following complement is very close. Therefore the verbal complementations to the right of the verb are not of an equal status: one of them is a part of a phraseme whereas the second one is a usual complementation externally connected with the whole verbal phraseme.

This type of the violation of the systemic order can be easily solved with the means used in Prague Dependency Treebank. In such cases, the complementations included in phrasemes are assigned with a label "compound phraseme". The excerpted sentences would not violate in this way the hypothesis of the systemic order if the list of phrasemes were enriched with the phrasemes found.

(7) [verb + PAT] + ADDR

*Opozici se nelíbí, že ODS [zadává úkoly.*PAT FOCUS*] státním úředníkům.*ADDR FOCUS

lit.: *Opposition doesn't like that ODS [charges with_tasks.*PAT FOCUS, ACC. PL.*] civil servants.*ADDR FOCUS, DAT. PL.

The opposition doesn't like that the Civic Democratic Party charges civil servants with tasks.

(8) [verb + PAT] + ACT

*Spolu s osmi ministry odjel přímo do Mardž Ujúnu, kde [má sídlo.*PAT FOCUS*] velení.*ACT FOCUS *SLA.*

lit.: *Together with eight ministers he_went directly to Marjayuon, where [has seat.*PAT FOCUS, ACC. SG.*] administration.*ACT FOCUS, NOM. SG. *of_SLA.*

He went together with eight ministers directly to Marjayuon, where the administration of South Lebanese Army seats.

4.4. A heavy complement is located at the end

In some sentences, the surface order of complementations which is contradictory to the systemic ordering is connected with the form of the complementations. The complements consisting of more words can be placed after formally shorter complements, regardless of the semantic structure of a sentence (functor, e.g. actor).

The surface word order differs in these cases from the underlying order, and this difference has to be taken into account in the description. The hypothesis of the systemic ordering concerns the underlying order, and the heavy complement movement should be taken into account in a specification of the conditions of the existence of differences between the underlying and the surface levels.

(9) PAT – ADDR

*Suchumsko-abchazský metropolita, gruzínský patriarcha David dnes zaslal poselství.*PAT _{FOCUS} *nejvyššímu představiteli muslimských věřících Kavkazu Allašukurovi Ulislamovi.*ADDR _{FOCUS}

  lit.: *Suchumi-Abchasian metropolitan, Georgian patriarch David today sent message.*PAT _{FOCUS, ACC. SG.} *to_the_highest representative of_Muslim believers of_Caucasus Allashucur Ulislam.*ADDR _{FOCUS, DAT. SG.}

  Suchumi-Abchasian metropolitan, Georgian patriarch David sent today a message to the highest representative of Caucasian Muslims, Allashucur Ulislam.

(10) PAT – ACT

*Zejména u začínajících podnikatelů má slušný ohlas.*PAT _{FOCUS} *kombinace.*ACT _{FOCUS} *tohoto účetního programu s počítačem MiniSys, která přijde na pouhých 25000 korun.*

  lit.: *Especially at beginning bussinessmen meets with a fair response.*PAT _{FOCUS, ACC. SG.} *combination.*ACT _{FOCUS, NOM. SG.} *of_this account program with computer MiniSys, which costs only 25 000 crowns.*

  Especially beginning bussinessmen appreciate the combination of this account program with the computer MiniSys, which costs 25 000 Czech crowns only.

4.5. A prepositional phrase is located after a non-prepositional phrase

Another deviation of the surface word order from the systemic ordering due to the form of a complementation can be observed in sentences with prepositional phrases. Analogously to the previous group of sentences (4.4), the longer or the more complicated complementation is placed after the shorter or the more elementary one. In this case, the prepositional phrase follows the non-prepositional phrase, irrespective of the distribution of functors (e.g. actor) and topic-focus articulation.

  This type of deviations from the systemic order should be explained similarly to the group of sentences in (4.4.): also in this case, the description needs to work with differences between underlying and surface forms of the complementations.

(11) PAT – ADDR

*Dvoutřetinová většina obyvatel ČR už nespojuje budoucnost.*PAT _{FOCUS} *se společným státem.*ADDR _{FOCUS.}

  lit.: *Two-third majority of_inhabitants of_Czech Republic anymore doesn´t connect future.*PAT _{FOCUS, ACC. SG.} *with common state.*ADDR _{FOCUS, SUBJUNCTION + INSTR. SG.}

  A two-third majority of inhabitants of the Czech Republic don´t connect their future with the common state (with Slovaks) anymore.

(12) PAT – ADDR

*Sami před sebou si tak vyřídili účty.*PAT _{FOCUS} *se svědomím.*ADDR _{FOCUS} *zatíženým tím, že válečný Slovenský štát poslal desetitisíce Židů do koncentračních táborů, řekl.*[3]

  lit.: *They for themselves so squared accounts.*PAT _{FOCUS, ACC. SG.} *with conscience.*ADDR _{FOCUS, SUBJUNCTION + INSTR. SG.} *burdened with that that wartime Slovac state sent tens_of_thousands Jews to concentration camps, he_said.*

  As he said, in this way, they have squared accounts with their conscience which was burdened with the fact that the Slovac state sent during the war tens of thousands Jews to concentration camps.

---

[3] There is another possible explanation of this example: the group of words *vyřídit účty* works here as a phraseme, similarly to the examples in 4.3, and it should be added to the list of phrasemes in the Prague Dependency Treebank.

4.6. Definition of the PAT functor

With the last group of sentences we found in the Prague Dependency Treebank as deviations from SO, the ground for the deviation is theory internal. This type of the deviation is characteristic for sentences with verbonominal predicates where nominal part of the predicate is usually located close to the verb, closer than other complementations.

The reason for the violation of the hypothesis lies in a difference between the conceptions of the patient functor in the general theoretical framework of FGD and in its part – the theory of the topic-focus articulation with the hypothesis of the systemic order. In FGD and subsequently in the Prague Dependency Treebank, the patient is understood as the functor of the second argument.[4] This concept is broad, it allows to interpret a nominal part of a verbonominal predicate (prototypically nominative) as a patient.[5] On the other hand, the theory of the topic-focus articulation and especially the hypothesis of the systemic order work implicitly with a notion of the patient in a narrow sense (prototypically accusative objects) – as we suppose, for historical reasons. This conception of the patient does not include the nominal part of a verbonominal predicate.

As for verbonominal predicates, the Prague Dependency Treebank has been annotated according to the general framework.[6] It follows from that that there are many sentences with verbonominal predicates where the nominal part of the predicate, annotated as patient, precedes another verbal complementation in focus, in contradiction to the systemic order. As for a possible solution, this type of sentences should be mentioned as a difference from the systemic order which is connected with the surface form of the complementation (similarly as in 4.4. and 4.5.).[7]

(12) PAT – ACT
*Je ale nutné.*PAT ₍FOCUS₎ *investovat.*ACT ₍FOCUS₎.
   lit.: *Is nevertheless necessary.*PAT ₍FOCUS, ADJ. NOM. SG. NEUTR.₎ *to_invest.*ACT ₍FOCUS, INF.₎
   Nevertheless, it is necessary to invest.

(13) PAT – ACT
*V postoji britské vlády se ale začaly projevovat.*PAT ₍FOCUS₎ *náznaky.*ACT ₍FOCUS₎ *změn.*
   lit.: *In attitude of_British government nevertheless started to_appear.*PAT ₍FOCUS, INF.₎ *signs.*ACT ₍FOCUS, NOM. PL.₎ *of_changes.*
   Nevertheless, signs of changes started to appear in the attitude of the British government.

5. Conclusion

It arises from the analysis of the data in Prague Dependency Treebank that there are several groups of sentences typically violating the hypothesis of the systemic order in Czech. In some cases, the deviation can be easily eliminated within the present framework of FGD: the complementation inversion in sentences with verbal phrasemes (4.3.) requires a more comprehensive list of verbal phrasemes only.

All the other cases of the deviation bring important arguments for an inclusion of some new points of view into the theory of the topic-focus articulation. It follows from the analysis that the hypothesis cannot work with the semantic structure of the sentence (functors) only. In some constructions, the form and the

---

[4] (Mikulová et al., 2005:465): "The PAT functor (Patient) is a functor used primarily for the second argument. In those cases when there is no argument shifting, the modification with the PAT functor refers to the affected object (in the broad sense of the word)."

[5] (Mikulová et al., 2005:467): "The PAT functor is also assigned to nodes representing the nominal part of a verbonominal predicate (e.g. *být hodný.*PAT (to be good))."

[6] (Mikulová et al., 2005:672): "Copula *být* has in its valency frame an Actor and a Patient. Here, the Patient comprises a broad spectrum of meanings which the verbonominal predicate can express… The congruity of the two constituents is not reflected in any specific way (other than by the selected valency frame)."

[7] It is possible to understand these constructions as phrasemes consisting of a verb and a patient as well (cf. examples in 4.3.).

lexical realization of complementations affect the surface word order, as in sentences with heavy complements (4.4.), with prepositional phrases (4.5.) and with the copula *být* (4.6.). These conditions should be formulated, tested on the data and added to the description of Czech word order.

In other constructions, the deviation from the systemic order seems to be caused by some specific features of the topic-focus articulation in Czech which we do not know exactly yet (the influence of the focus proper (4.1.) and of a contrastive contextually bound element (4.2.)). The research on the relations between the semantic structure, the topic-focus articulation and the word order in these constructions is crucial for further testing of the hypothesis of the systemic ordering. For these aims, the Prague Dependency Treebank provides a large amount of data which can be exploited.

6. Acknowledgements

References

Hajič, Jan et al. 2006. *Prague Dependency Treebank 2.0.* Linguistic Data Consortium, Philadelphia. (A demo version of the CD is available at http://ufal.mff.cuni.cz/pdt2.0)

Mikulová, Marie et al. 2005. *Annotation on the Tectogrammatical Layer in the Prague Dependency Treebank: Annotation Manual.* Universitas Carolina Pragensis, Prague.

Panevová, Jarmila. 1980. *Formy a funkce ve stavbě české věty.* Academia, Prague. *[Forms and Functions in the Czech Sentence Structure.]*

Preinhaelterová, Ludmila. 1997. Systemic ordering of complementations in English as tested with native speakers of British English. *Linguistica Pragensia,* pages 12–25.

Seidlová, Ivana. 1983. On the underlying order of cases (participants) and adverbials. *Prague Bulletin of Mathematic Linguistics,* 39, pages 53–64.

Sgall, Petr, Eva Hajičová and Eva Buráňová. 1980. *Aktuální členění věty v češtině.* Academia, Prague. *[Topic-Focus Articulation of Czech Sentences.]*

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

Sgall, Petr, W. Dressler, and M. Půček. 1995. Experimental research in systemic ordering. *Theoretical Linguistics,* 21, pages 197–239.

Sgall, Petr. 1967. *Generativní popis jazyka a česká deklinace.* Academia, Prague. *[Generative Description of the Language and Czech Declension]*