

Ordering of Adverbials of Time and Place in Grammars and in an Annotated English–Czech Parallel Corpus

Eva Hajičová, Jiří Mírovský, Kateřina Rysová

Charles University

Faculty of Mathematics and Physics

Institute of Formal and Applied Linguistics

[hajicova|mirovsky|rysova]@ufal.mff.cuni.cz

Abstract

The data from a parallel annotated English–Czech corpus serve for testing the general issue of the variability of the mutual position of LOC and TWHEN in Czech and English (Sect. 4.1) and for the analysis of the relation between information structure and the given order in the two languages (Sect. 4.2). The most relevant and innovative results in the investigation, namely the cases where the position of TWHEN and LOC differs in Czech and English in that the same modification is placed in Topic in the sentence in one language and in Focus in the corresponding sentence in the other are presented in Sect. 4.3.

1 Motivation and Research Question

In the early days of a massive entrance of corpus linguistics on the linguistic scene, C. J. Fillmore, in an attempt to characterize his own research position, compares two kinds of linguists: an armchair linguist and a corpus linguist. Fillmore (1992, 35) says: “Armchair linguist sits in his armchair, with his eyes closed and his hands clasped behind his back, once in a while, opens his eyes and shouts: Wow, what a neat fact”, while “Corpus linguist: has all of the primary facts he needs in the form of a corpus of approximately one zillion running words and he sees his job as that of deriving secondary facts from his primary facts.” And he concludes: “... the two kinds of linguists need each other. Or better, that the two kinds of linguists, wherever possible, should exist in the same body”. As for himself, he claims to be “an armchair linguist who refuses to give up his old ways but who finds profit in being a consumer of some of the resources that corpus linguists have created”.

In the era (and in the context) of treebanking, one can consider an armchair linguist to be a theoretically minded linguist and a corpus to be an annotated corpus in the form of treebanks, and it is in this sense that we have formulated our particular research question. The phenomenon under investigation is the *relation of word order and information structure*, the particular cases are *temporal and local modifications* of predicates and the data come from a *parallel English–Czech annotated corpus* (treebank).

The task we have faced is complicated by two facts: first, the information on structure is a very complex phenomenon and different approaches to its treatment have been proposed in theoretical literature since the pioneering studies by Czech scholars in the first half of the last century followed by such prominent linguists and semanticists as M.A.K. Halliday, B.H. Partee, M. Rooth, M. Krifka, E.F. Prince, K. Lambrecht, M. Steedman, E. Vallduví and E. Engdahl, to name just a few, and, second, it is hard to assess this phenomenon, so that annotation of information structure is very tricky (cf. Cook and Bildhouer, 2011) and therefore has to be carefully checked.

2 State of the Art

Though English representative grammars do not provide a systematic and comprehensive information on a possible variability of word order in English (which is quite understandable due to the predominance of grammatical factor determining the English SVO word order), it is somehow taken for granted, esp. in teaching English as a second language, that the unmarked order is SVOMPT, that is to say that with adverbials placed after the Object, Manner precedes Place and Place precedes Time. This more or less practical instruction is also reflected in Quirk et al. (1985, esp. parts 8.22–8.23): “Concerning adjuncts of the same grammatical class, subject to the stylistic and realizational factors already mentioned, will have their sequence determined by semantics and will normally appear in the order: process – space – time” (p. 650) giving examples such as *He worked at home that day.* or *The plane arrived uneventfully at Honolulu by midnight.* The authors continue: “Thus within the same class of adjuncts, those concerned with time are seen to be rather peripheral and this explains the case with which they can be moved to I (= initial position, EH): *By midnight, the plane arrived uneventfully at Honolulu.*” In the part on the relative positions of adjuncts (Chapter 8.87, pp. 565 ff.) the authors specify the order as respect – process – space – time – contingency, with two restrictions influenced by the information focus and the form of realization.

Leech and Svartvik (1994) mention the issues relevant for our investigation only briefly in the part on the position of adverbials (pp. 226–231) saying (p. 226) that “the place of an adverbial depends partly on its structure (whether it is an adverb, a prepositional phrase or clause, etc.), partly on its meaning (whether it denotes time, place, manner, degree, etc.). End-focus and end-weight also play a part.” The above-mentioned SVOMPT rule obtains here the following form: “When more than one of the main classes of adverbials occur in end-position, the normal order is manner/means/instrument + place + time.” The authors also take into account the influence of the form and the overall structure of the sentence, e.g. the fact that some adverbials which normally have an end-position can be in the front-position to avoid too many adverbials at the end of the sentence: *The whole morning he was working on his speech in the office.*

As for Czech, the relative freedom of surface word order makes it necessary to look for other than grammatical factors as determinants of the linear ordering of words in the sentence, the information structure being one of the main. In Vol. 3 of the representative Czech grammar *Mluvnice češtiny* (1987, p. 602) a “basic word order” is postulated, which is considered to be semantically based, reflecting the degrees of the so-called communicative dynamism (CD) as defined by the Czech anglicist Jan Firbas.¹ This basic word order may be influenced by the grammatical structure of the sentence, by its rhythmical structure and, marginally, by the size of the sentence elements in question. In the theory of information structure we subscribe to (the so-called topic–focus articulation, TFA, see e.g. Sgall et al., 1973; 1980; 1986) two orderings are postulated: one reflected in the surface shape of the sentence (surface word order) and the so-called underlying (deep) word order in the underlying (tectogrammatical) structure of the sentence. The underlying word order is semantically determined (and relevant), it reflects the TFA of the sentence and its counterpart in the surface is influenced, in addition to the TFA factors, by prosody, the overall structure of the sentence (e.g. the complexity of the structure), etc. One of the important notions introduced is the so-called systemic ordering (SO) as the order of verb modifications in the Focus part (see e.g. Sgall et al., 1980). The hypothesized order of main verb modifications is as follows: Actor – Temp – Cause – Regard – Aim – Manner – Accompaniment – Locative – Means – Addressee – Patient – Effect. The notion of SO in Focus is supposed to be universal, but the concrete order of modifications

¹ Cf. Firbas (1992, p. 105) “... the degree of CD carried by a linguistic element is the relative informational (communicative) value the element acquires in the development of the communication.”

may differ from language to language and has been already tested for some of them, see e.g. for German Sgall et al. (1995), for English Preinhaelterová (1997), for Czech Rysová (2014).

3 Methodology and Data

Our research question concerns the position of temporal and local modifications of predicates in Czech and English and the relation of this position to the information structure. The data come from a parallel English–Czech annotated corpus PCEDT (Hajič et al., 2012), which is a mostly manually annotated parallel corpus of English and Czech texts with almost 50 thousand sentences for each part. The E. part contains the Wall Street Journal section of the Penn Treebank (Marcus et al., 1993), along with the original phrase-structure analysis and a newly added dependency-based deep structure syntactic analysis (tectogrammatics). The Cz. part consists of manual translations of the original texts, along with their surface and deep syntactic analyses, automatically parsed and manually checked. We have analyzed the corpus findings and compared the results with claims made by existing representative grammars and other relevant studies and have tried to draw attention to contextual and other factors that play decisive role in the surface ordering of temporal (TWHEN) and locative (LOC) modifications. In doing so, we had in mind two limitations: the corpus data belong to the journalistic genre in which the TFA is not that clear as in other genres, and the translated sentences may be inclined to follow automatically the original order.

4 Queries and Corpus Findings

We have carried out a series of queries in which we were concerned with a general issue of variability of the mutual position of LOC and TWHEN in Czech and English (Sect. 4.1) and with the relation between TFA and the given order in the given languages (Sect. 4.2). The most relevant and innovative results in our investigation, namely the cases where the position of TWHEN and LOC differs in Czech and English in that the same modification is placed in Topic in the sentence in one language and in Focus in the corresponding sentence in the other are presented in Sect. 4.3.

4.1 Variability of the position of TWHEN and LOC

We have searched in the parallel corpus for cases with the Predicate as the root of the tree (excluding thus coordinated sentences) in which both TWHEN and LOC (occurring in the same tree) depend on the same Predicate. This search was carried out in the whole PCEDT, i.e. in the total of 39507 sentences with the Predicate as the root of the tree. The cases relevant for this step amount to 0.96% of the corpus. The results of our search are summarized in Table 1, where the E || Cz column refers to the number of cases in which the positions in Czech and English are the same.

It should be emphasized that the figures in Table 1 do not take into account the position of the modifications be it in the Topic or in the Focus, they just reflect the mutual positions of these modifications in the sentences in which both of them occur. The figures indicate that both orders are possible both in English and in Czech, and that in English the orders are relatively balanced (190 to 191), while in Czech the more frequent order is that of TWHEN before LOC (278 times when compared to LOC before TWHEN occurring 103 times).

	E Cz	E / Cz	Total E / Cz
LOC < TWHEN	85	105/18	190/103
TWHEN < LOC	173	18/105	191/278
Total	258	123	381

Table 1: The relative position of TWHEN and LOC in English and Czech in PCEDT

4.2 The relative position of TWHEN and LOC in the Focus part of the sentence

4.2.1 In the next step, we have taken into account the assumed division of the sentence into Topic and Focus and looked for cases in which both TWHEN and LOC were in the Focus part. The reasons why we have concentrated on the Focus part of the sentence, are twofold: first, and most importantly, we wanted to check whether and under which conditions the hypothesis of the above mentioned SO in the Focus is valid, both for English and for Czech, and, second, in this way, we could also check the before-mentioned general English word order “rule” SVOMPT, which indicates the order of Time after Place in the post-verbal position; with certain simplifications the post-verbal position may be considered to function as the Focus of the sentence.

We have tried first to search in that part of the PCEDT in which the sentences were annotated also as for their Topic–Focus articulation (3857 sentences), but the number of cases in which both TWHEN and LOC occurred in the same sentence in the relevant positions both in English and in Czech was very low (34 instances). Therefore we have decided to approximate the division into Topic and Focus as the position before (Topic) and after (Focus) the Predicate² and to carry out the search in the whole of PCEDT (on sentences with Predicate as the root of the tree), separately for English and for Czech.

	E.	E. after manual inspection	Cz.
TWHEN < LOC	129	103	164
LOC < TWHEN	202	130	90
TOTAL	331	233	254

Table 2: The occurrence of orderings of TWHEN and LOC in Focus in E. and in Cz. in PCEDT

The total number of sentences checked was 42717 for English and 39507 for Czech; the difference follows from the fact that there exist cases where one of the modifications is not realized by a separate sentence element.³ The results are given in Table 2.

The data obtained have made it possible to check the validity of the assumed so-called systemic ordering. The first attempt at such a verification for Czech was carried out by Rysová (2014) analyzing the data from the Prague Dependency Treebank 2.0 (PDT).⁴ The relevant figures in her Tables 6.1 (p. 77) and 6.10 (p.96) are summarized below in Table 3:

	Total number of occurrences	occurring in F	%
TWHEN	14552	4623	32
LOC	16948	10081	59
LOC < TWHEN		72	
TWHEN < LOC		332	

Table 3: The frequency of TWHEN and LOC (expressed by non-sentential elements) and their ordering in Focus in Czech in PDT 2.0 according to Rysová (2014)

² Such an approximation is based on the hypothesis common in many studies of information structure that the verb in principle stands on the boundary between the Topic and the Focus, cf. the notion of transition in Firbas (1992) and the analyses of Czech in Sgall et al. (1980) and Uhlířová (1974; 1987).

³ Comparing the English sentences containing LOC and TWHEN in PCEDT (without coordinated main predicates) with their Czech counterparts, LOC is missing in Cz. in 192 cases, and TWHEN is missing in Cz. in 88 cases. The difference is a consequence of several facts: the given modification in one language is translated by means of a different type of modification, a coordination structure is used in one language and not in the other, the given modification is understood as dependent on Noun rather than on a PRED, the dependency relations were understood differently, or a different structure is used in the translation.

⁴ The PDT (see the most recent version Hajič et al., 2018), contains approx. 50 thousand sentences of Czech journalistic texts annotated manually on several layers (morphology, surface and deep syntax) and contains also annotation of the topic-focus articulation of the sentences.

Rysová's results demonstrate that the data of PDT 2.0 support the SO as TWHEN < LOC; she also gives an explanation of the cases that do not correspond to this hypothesized order. Her observation is supported by the PCEDT data (see Table 2), though not so convincingly, which may be explained by the fact that the PCEDT data are translations and as such may mimicry to a considerable extent the E. order.

For English, our "raw" data indicate a different situation: TWHEN < LOC = 129 which is less than LOC < TWHEN = 202. However, after a manual inspection resulting in filtering out cases where the given modification, though placed after the verb, has to be characterized as contextually bound,⁵ i.e. belonging to the Topic part of the sentence, the figures attested were 103 for the TWHEN < LOC order, and 130 LOC < TWHEN order, which means that the preference for LOC < TWHEN is not so striking.

4.2.2 Let us first examine the examples of the TWHEN < LOC order, i.e. the order hypothesized by SO but counter to the assumed SVOMPT order. In 3 cases a decisive role was played by the form of the LOC modification as a clause (1).

(1) *Researchers began using the drug in February. TWHEN on patients. LOC who had received kidney, liver, heart and pancreas transplants.*

In the remaining 100 cases the LOC modification can be supposed to exemplify the order as predicted by SO. In most of them, both modifications are short (or of a comparable length) so that the "weight" criterion cannot be applied, see (2).

(2) *A volcano will erupt next month. TWHEN on the fabled Strip. LOC: a 60-foot mountain spewing smoke and flame every five minutes.*

With some examples, the TWHEN modification is closely related to the extralinguistic context (e.g. *today*) so that it can be understood as contextually bound and belonging to the Topic (3), though a different interpretation is also possible because in the preceding co-text District Court in Philadelphia is mentioned.

(3) *The trial begins today. TWHEN in federal court. LOC in Philadelphia. LOC.*

4.2.3 As for the LOC < TWHEN order, i.e. the order counter to the SO but in concord with the assumed SVOMPT order, we have again put aside examples in which TWHEN was expressed by a clause, which certainly had an impact on its end-position. This group was much larger than in the previous case, namely there were 48 examples in which the TWHEN modification was expressed by a clause, see (4):

(4) *Judy and I were in our back yard. LOC when the lawn started rolling like ocean waves. TWHEN*

The rest of the examples (82 sentences) mostly include the two modifications expressed by noun groups of a similar length (5), with an exception of some cases where the weight was a decisive factor (6).⁶

(5) *Mr. Guber got his start in the movie business at Columbia. LOC two decades. TWHEN ago.*

(6) *WASHINGTON lies low. LOC after the stock market's roller-coaster ride. TWHEN.*

⁵ In the TFA theory, on which the TFA annotation is based (see e.g. Sgall et al., 1986), contextual boundness is a primary notion interpreted as follows: A contextually bound node represents an item presented by the speaker as referring to an entity assumed to be easily accessible by the hearer(s), i.e. more or less predictable, readily available to the hearers in their memory. Each element of the underlying dependency tree of a given sentence is assigned one of the values of the TFA attribute, namely cb (contextually bound non-contrastive), c (contextually bound contrastive) or nb (contextually non-bound).

⁶ As remarked by one of the reviewers, "lie low" may be understood rather as an idiomatic expression.

4.2.4 We have also made a random inspection for particular cases where the parallel English and Czech sentences differed in the ordering of the two modifications. Interestingly enough, there are cases for which we have not found any reason why this was so, except for the “different ordering principles” (7).

(7) E.: *The company was founded in Sacramento. LOC in 1929. TWHEN by two brothers, Ralph and Walter Merksamer, who operated as DeVon's Jewelers.*

Cz.: *Společnost založili v roce 1929. TWHEN v Sacramentu. LOC bratři Ralph a Walter Merksamerovi pod jménem DeVon's Jewelers.*

However, having in mind that our parallel corpus was composed of translations from English to Czech, there was no surprise that the “principle ordering” in the target language was not obeyed and the Czech translation copied the order in E., see (8):

(8) E.: *Mr. Guber got his start in the movie business at Columbia. LOC two decades ago. TWHEN.*

Cz.: *Guber začínal ve filmové branži v Columbia. LOC před dvěma desítkami let. TWHEN.*

To sum up, while the SO for Cz. has been supported by both the PDT and the PCEDT data, the data for E. provide a slight support for the SVOMPT order.

4.3 Differences between Czech and English in the placement of TWHEN or LOC in the Topic and in the Focus

Most interesting for our study are the cases, where the two languages studied differ in the placement of the modifications TWHEN or LOC in the Topic in one language and in the Focus part of the same sentence in the other. In order to get a richer sample of examples, we have searched in the whole of PCEDT and we have again approximated the division into Topic and Focus by the position of these modifications before (Topic) and after (Focus) the main verb (PRED). We have at our disposal the samples in Table 4.

	TWHEN	LOC
In E. before PRED, in Cz. after PRED	233	67
In E. after PRED, in Cz. before PRED	765	271
TOTAL different order in E. and Cz.	998	338

Table 4: The position of TWHEN and LOC with respect to the Predicate in English compared to Czech

4.3.1 The position of TWHEN

We have randomly chosen a sample of 100 E. sentences and their Cz. counterparts from each of the sets (out of 233 and 765 examples, respectively) and analyzed them, also with regard to the previous context. The following observations seem to hold:⁷

A. TWHEN after the Predicate in English and before the Predicate in Czech

(i) Typically, TWHEN is expressed in E. by a short adverb (*-ly* adverb, *yesterday*, ...) and is placed next to the Predicate. In such a case, this post-verbal element may be considered to be a part of Topic also in E.

(9) E.: *In national over-the-counter trading, the company closed yesterday at \$23.25 a share.*

Cz.: *Při celostátním mimoburzovním obchodování společnost včera uzavřela na 23.25.*

(ii) TWHEN is expressed in E. by a short adverb and placed at the end of the sentence, but (presumably) this adverb does not carry the intonation centre; these examples, if analyzed properly with regard to Top-

⁷ In the examples, the relevant elements are underlined.

ic and Focus rather than with regard to its pre- or post-verbal position, would not represent instances of differences we are looking for (10).

(10) E.: *Democrats had been negotiating with some Republican congressional leaders on a compromise lately.*

Cz.: *V poslední době vyjednávali demokraté s některými čelními republikánskými představiteli Kongresu o kompromisu.*

(iii) In E., the position of TWHEN at the end of the sentence (i.e. in the prototypical position of Focus) is due to the weight of the element, being a prepositional phrase or a whole dependent clause (11).

(11) E.: *The shares traded at about A\$ 1.50 in March, when the plan to acquire MGM/UA was announced.*

Cz.: *V březnu, kdy byl plán na převzetí společnosti MGM/UA oznámen, se akcie obchodovaly kolem 1,50 australského dolaru.*

(iv) Nevertheless there was a considerable number of “true” examples where the E. sentence differed from its Cz. equivalent in the placement of the TWHEN modification in the Topic vs. the Focus part (12), (13):

(12) E.: *Coke introduced a caffeine-free sugared cola based on its original formula in 1983.*

Cz.: *Coke v roce 1983 uvedla na trh bezkofeinovou slazenou kolu založenou na původní receptuře.*

(13) E.: *But losers were spread in a broad range by the end of the session.*

Cz.: *Ale koncem burzovního dne se rozšířily řady těch, co ztratili.*

For some of these cases, as (14), the initial position of TWHEN in Cz. may be interpreted as a contrastive Topic: it is still (a part of) Topic, the sentence being “about” it, but the contrastive character of this element makes it comparable with Focus (which, as a choice of alternatives, always has a contrastive character).

(14) E.: *But we're ... going to be in the exact same situation next year.*

Cz.: *Ale příští rok budeme... v naprosto stejné situaci.*

B. TWHEN before the Predicate in English and after the Predicate in Czech

(i) A tendency observed by Czech grammars was attested in our data, to place the Predicate into the second position of the Cz. sentence, which has led to the placement of the TWHEN modification after the verb also in case in which it was an indisputable element of the Topic of the sentence (15):

(15) E.: *A year earlier, Nationwide Health earned.PRED \$2.4 million, or 29 cents a share.*

Cz.: *Výnosy společnosti Nationwide Health činily.PRED v loňském roce 2.4 milionu dolarů, neboli 29 centů na akcii.*

(ii) The difference in the placement of the TWHEN modification is due to the preferred position of short adverbs in E. (16):

(16) E.: *The utility company currently has about 82.1 million shares outstanding.*

Cz.: *Tento podnik veřejných služeb má v současné době v oběhu 82.1 milionu akcií.*

(iii) However, even in this group, quite clear examples are found testifying the difference in Topic and Focus in E. and in Cz.; in some cases, the initial position should be understood as a contrastive Topic (17), see Quirk et al. (1985) where fronting is mentioned as a regular means for emphasizing a contrastive Topic.

(17) E.: *Only twice since the 1960s has annual gross domestic product growth here fallen below 5% for two or more consecutive years.*

Cz.: *Roční nárůst hrubého domácího produktu zde spadl pod 5 % během dvou nebo více po sobě jdoucích let pouze dvakrát od šedesátých let.*

4.3.2 The position of LOC

We have again randomly chosen 100 sentences from the set of LOC after PRED in E. and we have analyzed all the sentences in the set of LOC before PRED in E. (i.e. the total of 67 sentences) taking into consideration also the previous context. The following observations seem to hold:

A. LOC after the Predicate in English and before the Predicate in Czech

(i) As has been mentioned above in our discussion on the sentences with TWHEN, the position of a modification close to the Predicate may be considered as a part of Topic or alternatively as a part of Focus, as the example below demonstrates:

(18) E.: *The two boards said.PRED in a joint statement that the proposed merger agreement was considered in separate board meetings in Oslo Monday.*

Cz.: *Obě správní rady ve společném prohlášení uvedly.PRED, že navrhovaná dohoda o sloučení byla v pondělí posouzena na jednotlivých zasedáních správních rad v Oslu.*

(ii) The final position in E. need not be an indicator of the Focus position because the given element need not be a carrier of the intonation center; the prosodic factor is decisive here for the identification of Focus (19):

(19) E.: *Logic plays a minimal role here.*

Cz.: *Logika tady hraje minimální roli.*

(iii) A modification is placed at the end of the sentence in E. because of its weight, which does not necessarily mean that this modification is in Focus (20):

(20) E.: *The topic never comes up in ozone depletion "establishment" meetings, of which I have attended many.*

Cz.: *Toto téma se na „schvalovacích“ schůzích o ozónové díře, kterých jsem navštívil hodně, nikdy neujme.*

(iv) The placement of the modification is given by grammatical restrictions of word order in E., namely that subject should precede the verb (21); there belong also examples with *there*-construction (22):

(21) E.: *A tractor, his only mechanized equipment, stands in front of the pigsty.*

Cz.: *Před prasečím chlívem stojí traktor, jeho jediné mechanizované zařízení.*

(22) E.: *There was no new-issue activity in the derivative market.*

Cz.: *Na trhu odvozených cenných papírů nebyla vyvíjena žádná nová emisní aktivita.*

(v) The dependency relation is different in the original and in the translation: in E. “on television” depends on “events”, while in Cz. the LOC is understood as a modification of the verb (23).

(23) E.: *The Series typically is among the highest-rated sports events on television.*

Cz.: *V televizi světová série obvykle patří mezi nejvýše hodnocené sportovní události.*

(vi) However, similarly as is the case with the placement of the modification TWHEN, there was a considerable number of “true” examples where the original E. sentence differed from its Cz. equivalent in the placement of the LOC modification in the Topic vs. in the Focus part (24), (25).

(24) E.: *The citation was misstated in Friday's edition.*
Cz.: *V pátečním vydání byla tato citace uvedena chybně.*

(25) E.: *Each has an equal vote at the monthly meetings.*
Cz.: *Na měsíčních schůzích mají všichni stejný hlas.*

It is often the case that the preceding context helps to identify the Focus, but not necessarily so, as the following example demonstrates (26):

(26) E.: *The year was misstated in Friday's edition.*
Cz.: *V pátečním vydání byl rok uveden chybně.*

E. previous context: *QUANTUM CHEMICAL Corp.'s plant in Morris, Ill., is expected to resume production in early 1990.*

B. LOC before the Predicate in English and after the Predicate in Czech

The analysis of the examples with LOC before the Predicate in E. and after the predicate in Cz. has led to observations analogous to those mentioned in Sect 4.3.1 B above. Similarly as noted in 4.3.2 A, a tendency was also observed to place the Predicate into the second position of the Cz. Sentence, which has led to the post-verbal placement of the LOC modification also in case in which it was an indisputable element of the Topic of the sentence, see (27).

(27) E.: *In an interview, Pemberton Hutchinson, president and chief executive, cited several reasons for the improvement: higher employee productivity and “good natural conditions” in the mines, as well as lower costs for materials, administrative overhead and debt interest.*

Cz.: *Prezident a výkonný ředitel Pemberton Hutchinson jmenoval.PRED v rozhovoru několik důvodů zlepšení: vyšší produktivitu zaměstnanců a „dobré přírodní podmínky” v dolech, stejně jako nižší cenu materiálu, administrativní režii a úroky z úvěrů.*

Comparing the number of sentences in which the position of LOC with regard to the Topic and Focus position in Cz. and E. differed, it should be noted that in E., LOC occurred relatively much less frequently in the front position than in the Focus position (23% to 77%). Interestingly enough, almost the same proportion holds for TWHEN, which occurred in 20% in the front position and in 80% post-verbally. It seems that the final position of both of these modifications is the preferred one.

5 Summary

Our main concern has been the relation of word order and information structure in English and in Czech, in particular the mutual order of temporal and local modifications of predicates. We have put under scrutiny the data from the annotated parallel English–Czech treebank (PCEDT) and tested the variability of the order of the given types of modifications in general and two hypotheses on their preferential order in particular, namely the SVOMPT hypothesis for English and the so-called systemic ordering hypothesis for both languages. Our probe has demonstrated that corpus data offer much richer material to work with than an “arm-chair” linguist has ever had at her/his disposal but also that a careful manual check is necessary to obtain a reliable source for a detailed linguistic analysis that eventually may lead to some well-founded theoretical conclusions.

Acknowledgements

The authors are deeply indebted to Prof. Libuše Dušková, the leading Czech anglicist, for her observations and comments concerning the topic of this contribution.

The authors also gratefully acknowledge support from the Grant Agency of the Czech Republic (projects GA1703461S and GA17-06123S) and the Ministry of Education, Youth and Sports of the Czech Republic (project LM2015071). The research reported in the present contribution has been using language resources developed, stored and distributed by the LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (LM2015071 and OP VVV VI CZ.02.1.01/0.0/0.0/16 013/0001781).

References

- Cook, Philippa., Bildhauer Felix. (2011), Annotating Information Structure. The Case of “Topic”. In Dipper, S., Zinsmeister H. (eds.), *Beyond Semantics. Corpus-based Investigations of Pragmatic and Discourse Phenomena*. Bochum: Ruhr-Universität Bochum, 45–56.
- Fillmore, C. J. (1992), “Corpus linguistics” or “Computer-aided armchair linguistics”, In: *Directions in Corpus Linguistics*, Proceedings of Nobel Symposium 82 Stockholm, 4-8 August 1991, ed. Jan Svartvik, Mouton De Gruyter, Berlin New York, pp. 61–77.
- Firbas, Jan (1992), *Functional Sentence Perspective in Written and Spoken Communication*. Cambridge, Cambridge University Press.
- Hajič, Jan, Bejček Eduard, Bémová Alevtina et al. (2018), *Prague Dependency Treebank 3.5*. Data/Software. Institute of Formal and Applied Linguistics, LINDAT/CLARIN, Charles University. PID: <http://hdl.handle.net/11234/1-2621>.
- Hajič, Jan, Hajičová Eva, Panevová Jarmila et al. (2012), Announcing Prague Czech-English Dependency Treebank 2.0. In *Proceedings of the LREC 2012*, pp. 3153–3160.
- Leech, Geoffrey and Jan Svartvik, 1994. *A Communicative Grammar of English*. 2nd ed., London: Longman.
- Marcus, M. P., Marcinkiewicz, M. A. and B. Santorini (1983), Building a Large Annotated Corpus of English. The Penn Treebank. *Computational Linguistics* 19 (2), pp. 313–330.
- Mluvnice češtiny* 3 (1987), Prague, Academia.
- Preinhaelterova, Ludmila (1997), Systemic ordering of complementations in English. *Philologica Pragensia* 1997, pp. 12–25.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik (1985), *A Comprehensive Grammar of the English Language*, London and New York: Longman.
- Rysová, Kateřina (2014), *O slovosledu z komunikačního pohledu* [On Word Order from a Communicative Point of View], Studies in Computational and Theoretical Linguistics 12, ÚFAL Charles University, Prague.
- Sgall, Petr, Hajičová, Eva and Eva Benešová (1973). *Topic, Focus, and Generative Semantics*, Kronberg/Taunus: Scriptor.
- Sgall, Petr, Hajičová, Eva and Eva Buráňová (1980), *Aktuální členění věty v češtině*. [Topic-Focus Articulation of the Czech Sentence]. Prague: Academia.
- Sgall, Petr, Hajičová, Eva and Jarmila Panevová (1986). *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Dordrecht: Reidel, Prague: Academia.
- Sgall, Petr, Oskar Pfeiffer, W. U. Dressler and Milan Půček (1995), Experimental research on systemic ordering, *Theoretical Linguistics* 21: pp. 197–239.
- Uhlířová, Ludmila (1974), O vztahu sémantiky příslovecného určení k aktuálnímu členění. [On the relation of semantics of adverbials to the information structure]. *Slovo a slovesnost* 35, pp. 99–106.
- Uhlířová, Ludmila (1987). *Knížka o slovosledu* [A book on word order]. Prague.