



FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University



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, Prague

1. Motivation and Research Question

From theory to corpus annotation and back to theory (Ch. Fillmore)

Research question:

- Phenomenon under investigation: **relation of word order and information structure**
- Particular case: **temporal and spatial** modifications of verbs
- Data: **parallel English – Czech** annotated treebank (PCEDT)

Expected obstacles

The task is complicated (at least) by three facts:

- (i) **information structure** (IS): a **complex** phenomenon, different approaches
- (ii) **annotation** of IS is very **tricky** and therefore has to be carefully checked manually
- (iii) the PCEDT texts are **translations** so that the target Czech sentences may mimicry the source English sentences

Outline of the talk

1. **Treatments of word order** in representative grammars of Cz. and E.
2. **Methodology** and **Data**
3. **Queries** and **Results** obtained:
 - (i) variability of the **position of TWHEN and LOC** in general
 - (ii) relative position of TWHEN and LOC **in the Focus** part of the sentence
 - (iii) **differences** in the placement of TWHEN and LOC in the Topic and in the Focus
4. **Summary** and **Results**

1. Treatments of word order in representative grammars of Cz. and E.

English: WO grammaticalized -> grammars do not provide a systematic information

- Teaching E.: **SVOMPT** order assumed, see also Quirk et al. (1985, parts 8.22 and 8.23) -> **spatial** before **temporal**
- Important role: **end-focus** and **end-weight** (Leech and Svartvik 1994, 226-231)

Word order in Czech

Czech: other than grammatical factors: **semantically based**

One of the most important factors: information structure

Hypothesis of the so-called **systemic ordering** (SO) in the **Focus** part of the sentence

- Actor – **Temp** – Cause – Regard – Aim – Manner – Accompaniment – **Locative** – Means – Addressee – Patient – Effect

Systemic ordering: the notion is **universal**, but the concrete order of modifications may differ from language to language (already tested, e.g. for German Sgall et al. (1995), for English Preinhaelterová (1997), for Czech Rysová (2014))

2. Methodology and Data

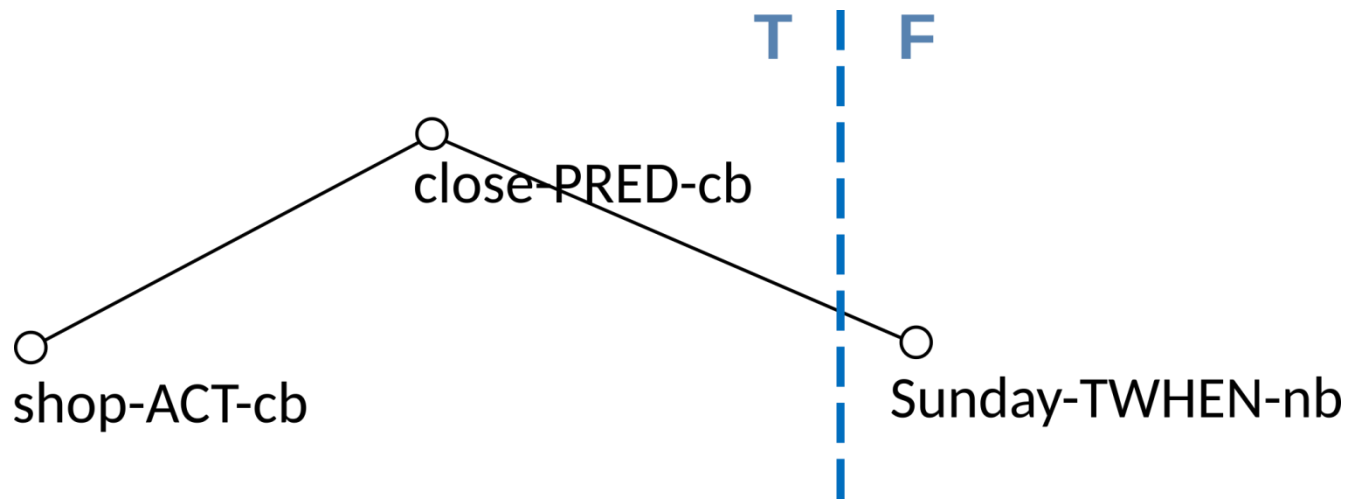
Data:

a **parallel English-Czech annotated** corpus PCEDT

- mostly manually annotated parallel corpus of English and Czech
- almost **50 000 sentences** for each part
- **English part**: the WSJ section of Penn Treebank, along with newly added dependency-based deep structure syntactic analysis
- **Czech part**: manual translations of the original texts, along with their surface and deep syntactic analyses, automatically parsed and manually checked.

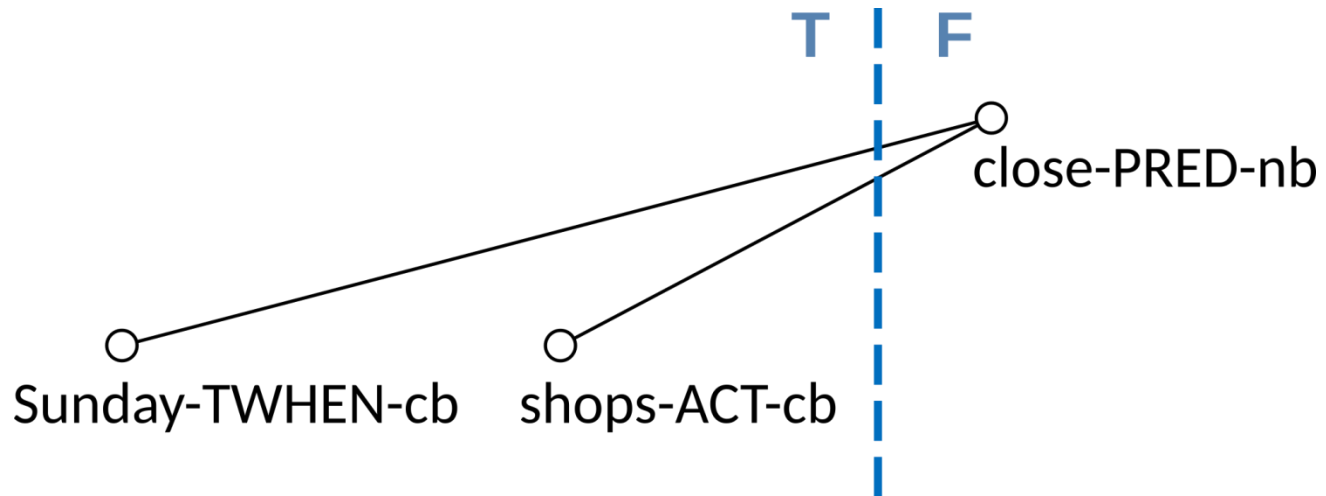
Annotation:

- **temporal** (TWHEN) and **spatial** (LOC) modifications
- **TFA attribute**: contextual boundness, algorithm for T/F dichotomy



(When do shops close?)

Shops close on Sundays.



(What about the shops on Sundays?)

On Sundays, shops close.

3. Queries and results obtained (i)

(i) **variability of the position of TWHEN and LOC in general**

Predicate: the root of the tree (ie. without coordination)

Dependents: both TWHEN and LOC occurring in the **same** tree

The search carried out in the whole PCEDT (**39507 sentences** with the Predicate as the root of the tree)

The cases relevant for this step: 0.96% of the corpus.

	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

3. Queries and results obtained (i)

(i) variability of the position of TWHEN and LOC in general

Predicate: the root of the tree (ie. without coordination)

Dependents: both TWHEN and LOC occurring in the **same** tree

The search carried out in the whole PCEDT (39507 sentences with the Predicate as the root of the tree)

The cases relevant for this step: 0.96% of the corpus.

	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

3. Queries and results obtained (ii)

- (ii) relative position of TWHEN and LOC in the Focus part of the sentence
 - (a) testing the hypothesis of SO in the Focus, both for English and for Czech,
 - (b) testing the English WO “rule“ SVOMPT: Time after Place in the post-verbal position

Two steps:

1. search in the part of the PCEDT with TFA annotation (3857 sentences):
both TWHEN and LOC occurred only in 34 instances
2. approximation of the division into Topic and Focus as the position before (Topic) and after (Focus) the Predicate -> the search in the whole of PCEDT:
based on the hypothesis the verb in principle stands on the boundary between T and F

Cf. the notion of transition in Firbas (1992, Functional Sentence Perspective in Written and Spoken) and the analyses of Czech in Sgall et al. (1980, Topic-Focus Articulation of the Czech Sentence) and Uhlířová (1974, On the relation of semantics of adverbials to the information structure; 1987, A book on word order).

3. Queries and results obtained (ii)

- total number of sentences checked: 42717 for English and 39507 for Czech
- reasons for the different numbers:
 - i. the given modification is translated by a different type
 - ii. a coordination structure
 - iii. the head: Verb vs. Noun
 - iv. a different structure is used in the translation

	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

3. Queries and results obtained (ii)

(a) Testing the hypothesis of **systemic ordering** in Focus

For **Czech**: Rysová (2014): data: Czech annotated PDT, support for TWHEN < LOC

	Occurrences in the PDT
TWHEN < LOC	332
LOC < TWHEN	72

PCEDT: not so convincing (164 vs. 90): explanation: not original data, but translations!

3. Queries and results obtained (ii)

(b) For **English**:

TWHEN < LOC: according to SO, counter to **SVOMPT**: 103 cases

The trial begins.PRED today.TWHEN in Federal Court.LOC in Philadelphia.LOC

LOC < TWHEN: according to **SVOMPT**, counter to SO: 130 cases

Mr. Guber got.PRED his start in the movie business at Columbia.LOC two decades.TWHEN ago.

Conclusion: the data for E. provide a slight support for the **SVOMPT** order

3. Queries and results obtained (iii)

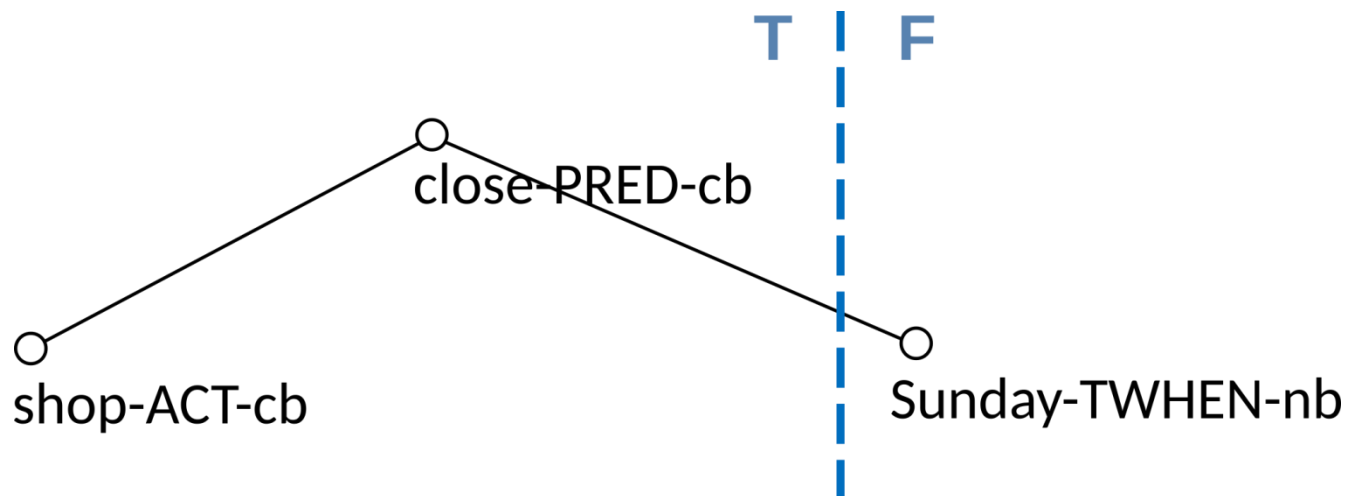
(iii) **differences** between Cz. and E. in the placement of TWHEN and LOC **in the Topic** in one language and **in the Focus** part of the same sentence in the other

= core of our study

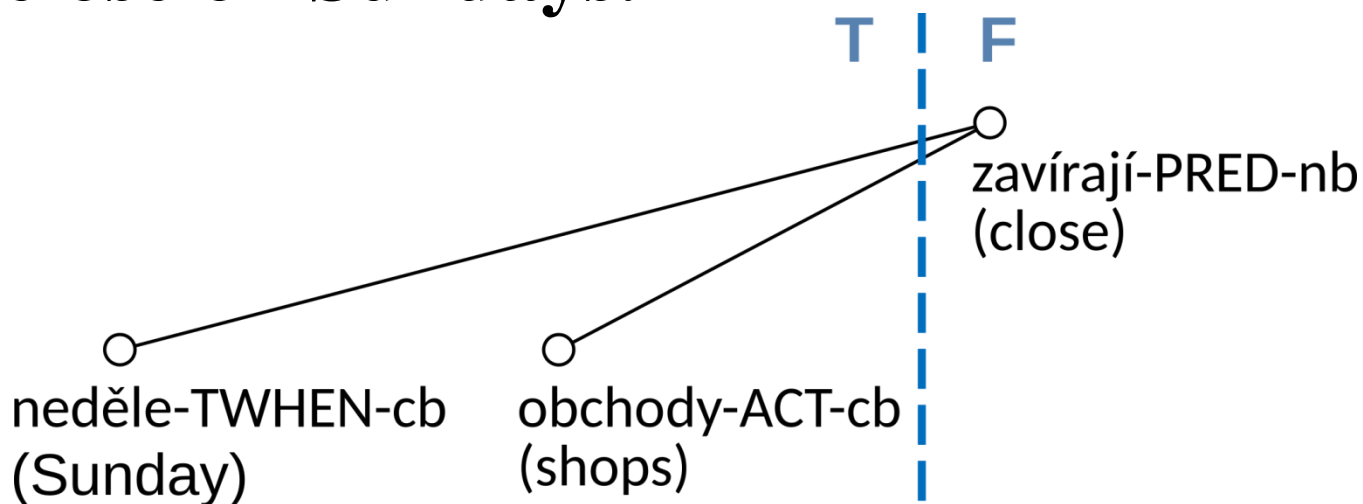
- to get a richer sample of examples: search in the **whole of PCEDT** with **approximation of the division** into Topic and Focus by the position of these modifications before (Topic) and after (Focus) the main verb (PRED).

	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



Shops close on Sundays.



**V neděli obchody zavírají.
[On Sundays shops close.]**

(1) The position of **TWHEN**

(1) The position of **TWHEN**:

TWHEN before and after PRED: sample of 100 **English** sentences and their **translations** from each set

(a) **E.: TWHEN > PRED, Cz: TWHEN < PRED**

(i) **short adverb** → Topic?

- E.: *In national over-the-counter trading, the company **closed.PRED yesterday** at \$23.25 a share.*
- Cz.: *Při celostátním mimoburzovním obchodování společnost **včera uzavřela.PRED** na 23.25.*

(ii) **short adverb** at the end, but **without IC** → Topic?

- E.: *Democrats had been negotiating.PRED with some Republican congressional leaders on a compromise **lately**.*
- Cz.: ***V poslední době** vyjednávali.PRED demokraté s některými čelními republikánskými představiteli Kongresu o kompromisu.*

(iii) **weight** of the final element:

- E.: *The shares traded.PRED 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.PRED kolem 1,50 australského dolaru.*

True differences

- E.: *Coke introduced.PRED a caffeine-free sugared cola based on its original formula in 1983.*
- Cz.: *Coke v roce 1983 uvedla.PRED na trh bezkofeinovou slazenou kolu založenou na původní receptuře.*
- E.: *But losers were spread.PRED in a broad range by the end of the session.*
- Cz.: *Ale koncem burzovního dne se rozšířily.PRED řady těch, co ztratili.*

a **contrastive Topic**? still (a part of) Topic, the sentence being "about" it, but the contrastive character of this element makes it comparable with Focus (which always has a contrastive character)

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

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

E.: TWHEN > PRED x Cz: TWHEN < PRED

E.: LOC > PRED x Cz.: LOC < PRED

E.: LOC < PRED x Cz.: LOC > PRED

→ similar possible explanations of the differences, but true differences there, too

The **preceding context** need not help to identify the Focus:

- E.: *The year was misstated. PRED in Friday's edition.*
- Cz.: *Vpátečním vydání byl rok uveden. PRED chybně.*
- E. previous context: *QUANTUM CHEMICAL Corp.'s plant in Morris, Ill., is expected to resume production in early 1990.*

4. Summary

Main objective: 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.

Data: the annotated parallel **English-Czech treebank** (PCEDT)

Queries: testing the variability of the order of the given types of modifications in general and two hypotheses on their preferential order

- (i) the **SVOMPT hypothesis** for English
- (ii) the so-called **systemic ordering hypothesis** for both languages.

Results

- (i) the data for **English** provide a **slight support** for the **SVOMPT** order
- the **final position** of both of these modifications seems to be the **preferred** one in English
- (ii) for **Czech**: the **systemic ordering** of Time < Place slightly confirmed but not so convincingly as in PDT original texts(164 vs. 90): explanation: not original data, but translations
- (iii) examples of **true differences** in the topic vs. focus position of TWHEN and LOC are rather rare but **DO exist**



THANK YOU FOR YOUR ATTENTION!

Questions?

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 GA17-03461S and GA17-06123S) and the Ministry of Education, Youth and Sports of the Czech Republic (project LM201507).

This work has been supported and using language resources and tools 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).

(b) E.: TWHEN > PRED, Cz: TWHEN < PRED

(i) short adverb → Topic?

- 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 placement of the TWHEN modification is due to the preferred position of short adverbs in E.:

- E.: The utility company *currently* has.PRED about 82.1 million shares outstanding.
- Cz.: Tento podnik veřejných služeb má.PRED v současné době v oběhu 82.1 milionu akcií.

True differences?

(iii) quite **clear examples** of the difference in Topic and Focus in E. and in Cz.; in some cases, the initial position should be understood as **contrastive Topic**:

- E.: *Only twice since the 1960s has annual gross domestic product growth here fallen. PRED below 5% for two or more consecutive years.*
- Cz.: *Roční nárůst hrubého domácího produktu zde spadl. PRED pod 5 % během dvou nebo více po sobě jdoucích let pouze dvakrát od šedesátých let*

The position of LOC

a) **E. LOC > PRED** x **Cz. LOC < PRED**

(i) LOC **close to verb** may be Topic or Focus

- 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) LOC is **not the IC** -> not an end-focus

- E.: *Logic plays.PRED a minimal role here.*
- Cz.: *Logika tady hraje.PRED minimální roli.*

E. LOC > PRED x Cz. LOC < PRED (Cont.)

(iii) E.: the decisive factor is the **weight** rather than Focus:

- E.: *The topic never comes up.* PRED *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.* PRED

(iv) **grammatical word order** in E., namely that subject should precede the verb

- E.: *A tractor, his only mechanized equipment, stands.* PRED *in front of the pigsty.*
- Cz.: *Před prasečím chlívem stojí.* PRED *traktor, jeho jediné mechanizované zařízení.*

.

True differences

“true“ examples of difference in placement of LOC in T or F

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

The preceding context need not help to identify the Focus:

- E.: *The year was misstated.PRED in Friday's edition.*
- Cz.: V pátečním vydání byl rok uveden.PRED chybně.
- E. previous context: *QUANTUM CHEMICAL Corp.'s plant in Morris, Ill., is expected to resume production in early 1990.*

(b) E.: **LOC < PRED, Cz. LOC > PRED**

observations **analogous** to those for TWHEN

a tendency to place PRED into the **2nd position** in **Cz.** -> post-verbal placement of the LOC modification -> indisputable **element of the Topic** of the sentence:

- E.: *In an interview, Pemberton Hutchinson, president and chief executive, cited PRED 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.*

- Observation for English:
- LOC occurred relatively much less frequently in the front position than in the Focus position (23% to 77%)
- almost the same proportion holds for TWHEN, which occurred in 20% in the front position and in 80% post-verbally
- -> the final position of both of these modifications seems to be the preferred one in English