FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

ÚFAL

LINDAT
CLARIN

Syntax – Semantics Interface: A Plea for a Deep Dependency Sentence Structure

Eva Hajičová

Charles University

Prague, Czech Republic

In collaboration with

Václava Kettnerová, Veronika Kolářová, Markéta Lopatková, Jarmila Panevová, Dan Zeman and Jan Hajič

(and with technical support of Jiří Mírovský)

Supported by the LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (project LM2015071) and by the projects No. GA17-07313S and No. GA17-06123S of the Grant Agency of the Czech Republic.

MOTIVATION

two central objectives: arguments for Deep and Dependency

(a) the argumentation will be based on linguistic considerations

(b) the linguistic background – Praguian:
the structuralist tradition of the Prague School
the formal framework: Functional Generative Description
the experience with the building of the Prague Dependency Treebank

(c) collective work - the arguments are not novel but based on
collected experience with the study and application of the deep syntactic dependency relations in the description of language;
basic material: Czech, but comparative aspects taken into account as well

OUTLINE

1. Motivating considerations
2. Deep (underlying, tectogrammatical, …)
   - 2.1 Linguistic meaning vs. cognitive content
   - 2.2 Synonymy and Ambiguity
   - 2.3 Multilevel treatment of syntax: The case of PDT
3. Dependency and valency
   - 3.1 Historical excursion
   - 3.2 Arguments and adjuncts
4. Grammar and lexicon
5. Challenges
   - 5.1 Non-projectivity
   - 5.2 Coordination and apposition
   - 5.3 Reconstruction of surface deletions

LOVÍ	TLOUŠTĚ	NA	VIŠNI
CATCH(ES)	CHUB	ON	MORELLO

- loví “catch” he/she/it/they -catch
- tlouště “chub” Gen/Acc sg, Nom/Acc pl
- na “on”
- višni Dat/Loc sg of “višeň”
  = “morello” / “morello-tree”
- structural interpretations:
  Subj – Verb – Object – Loc
  Is Subject or Object on the tree?
  Subj – Verb – Object – Instrument (?Manner)
  = morello as a bait put on a hook to catch fish

2. DEEP (UNDERLYING, TECTOGRAMMATICAL, SEMANTIC …)

"To understand the ability of natural languages to serve as instrument to the communication of thoughts and ideas we must understand what it is that permits those who speak them consistently to connect the right sounds with the right meanings" (Katz 1966, p. 100)

"meanings"?

Task of linguistics: to specify the relation between the outer shape of sentences and their semantic representations (meanings) => how "deep"?

Katz (1966), The Philosophy of Language

2.1 (Cognitive) Content and (Linguistic) Meaning

the borderlines between **linguistic meaning** and **cognitive content**:

Illustration: semantic relationships between a verb and its complementations:

		Tesnière (1959)	Fillmore (1968)
The smoke rose.	1st actant		Objective
John likes fish.			Dative
A hammer broke the window.			Instrument
The tree was struck by lightning.			Locative

Sgall & Hajičová (1970), A Functional Generative Description;
Tesnière (1959), Élements de syntaxe structurale;
Fillmore (1968), The Case for Case; Halliday (1967), Notes on Transitivity

2.2 SYNONYMY AND AMBIGUITY

Two relations important with regard to the specification of the level of meaning;

- (a) synonymy
- (b) ambiguity

- Synonymous sentences: the same representation on the level of meaning, different forms
- Ambiguous sentences: different representations, same form

THE RELATION OF SYNONYMY

Synonymy:

two sentences differing in the given semantic opposition have the same truth conditions

= there does not exist a situation when one sentence would be true while the other sentence would not

a proof of non-existence is not possible

→ always has a nature of a hypothesis

but a useful criterion for interchangeability

Panevová (1980), Formy a funkce [Forms and Functions]

INTERCHANGEBILITY AS A CRITERION FOR SYNONYMY

Pavel sold Jirka a car.

↔ *Jirka bought a car from Pavel.*

... with enthusiasm: not synonymous

He cut salami into five pieces.

↔ *He cut five pieces from salami.*

... and some salami was still left: not synonymous

synonymous:

I promised I will do it in time. ↔ *I promised to do it in time.*

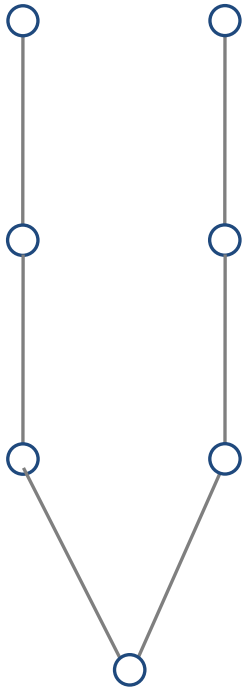
HOMONYMY (AMBIGUITY)

tectogrammatical

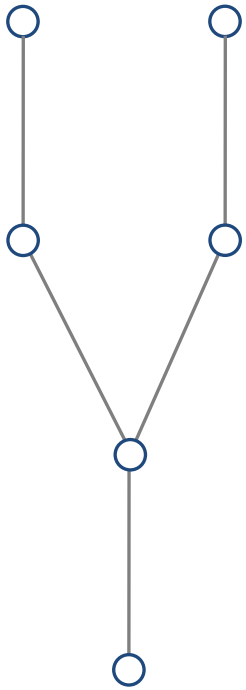
surface syntax

morphological

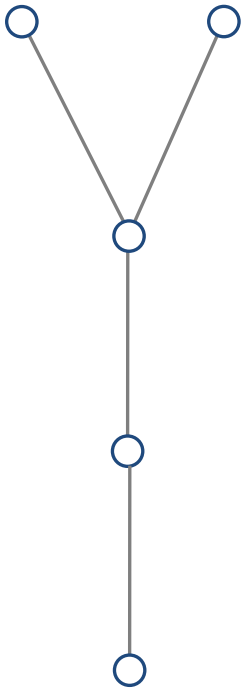
phonological



slepice
[hen/s]



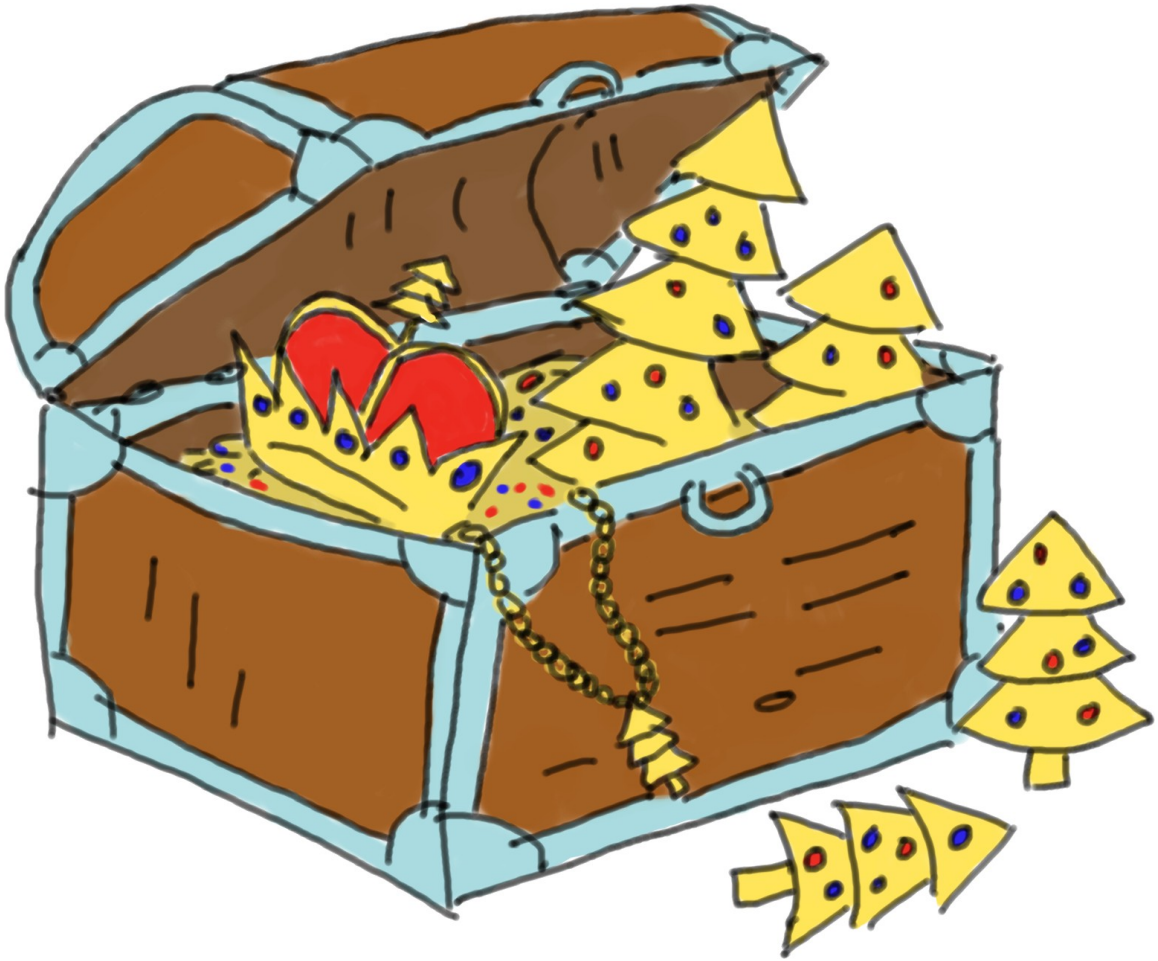
He saw a man
with a
telescope



criticism of
the delegate

Panevová (1980), Formy a funkce [Forms and Functions]

2.3. Multilevel Descritpion: The case of PDT



MULTILAYER ANNOTATION SCHEME OF PDT (1)

The inclusion of an underlying (deep) level into the theoretical description of a language ➔ postulation of a multilayer scheme

- The Prague Dependency treebank (PDT), v 3.0, Czech

- The Prague Czech – English Dependency Treebank, v 2.0, WSJ from Penn Treebank

- The Prague Dependency Treebank of Spoken Czech, v 1.0, conversations, interviews

- The "Faust" project treebank (Czech), small, user-generated content

- The Prague Arabic Dependency Treebank, sample annotated on TR

all (manually) annotated on the deep (tectogrammatical) layer and available from the LINDAT/CLARIN repository at http://lindat.cz

MULTILAYER ANNOTATION SCHEME OF PDT (2)

based on the **theory** of Functional Generative Description (formulated in the 1960s)

the "deep" syntactic **"tectogrammatical"** level: dependency
conceptually **separated** from the surface dependency structure and its annotation
full alignment between the elements (tree nodes) of both annotation layers
continuous Czech texts
manual annotation

Size: total number of documents annotated on all levels: 3,168,
= 49,442 sentences, 833,357 tokens

PDT versions 1.0 and 2.0 available from the Linguistic Data Consortium

Other additions (such as discourse annotation) in PDT 2.5 and in **PDT 3.0**, both available from the LINDAT/CLARIN repository (Bejček et al. 2013)

Hajič (1998), Building a Syntactically Annotated Corpus; Bejček et al. (2013), Prague Dependency Treebank 3.0; Hajič, Hajičová, Panevová et al. (2011), Prague Czech-English Dependency Treebank 2.0

LAYERS OF PDT ANNOTATION

(a) morphological layer: all elements (tokens) of the sentence get a lemma and a (disambiguated) morphological tag (= combination of morph. features)

(b) analytical layer:

a dependency tree, with dep. relations such as subject, object, adverbial

nodes for all and only lexically realized elements in the surface

(c) tectogrammatical layer: deep dependency syntactic relations

- nodes only for autonomous meaningful units ("content" words)
- function words (prepositions, conjunctions, auxiliary verbs etc.) not as separate nodes – "converted" to labels at the autonomous units

  WHY? parallel with inflectional endings, affixes ...
- (deep) dependency relations such as Actor, Patient, Addressee, and different kinds of circumstantial relations
- (deep) multiword expressions: subtrees *(higher education institution)*
- topic-focus articulation: contextual boundness
- coreference: grammatical, textual, bridging
- discourse relations: on top of the previous annotation

COMPLEX LABELS OF NODES

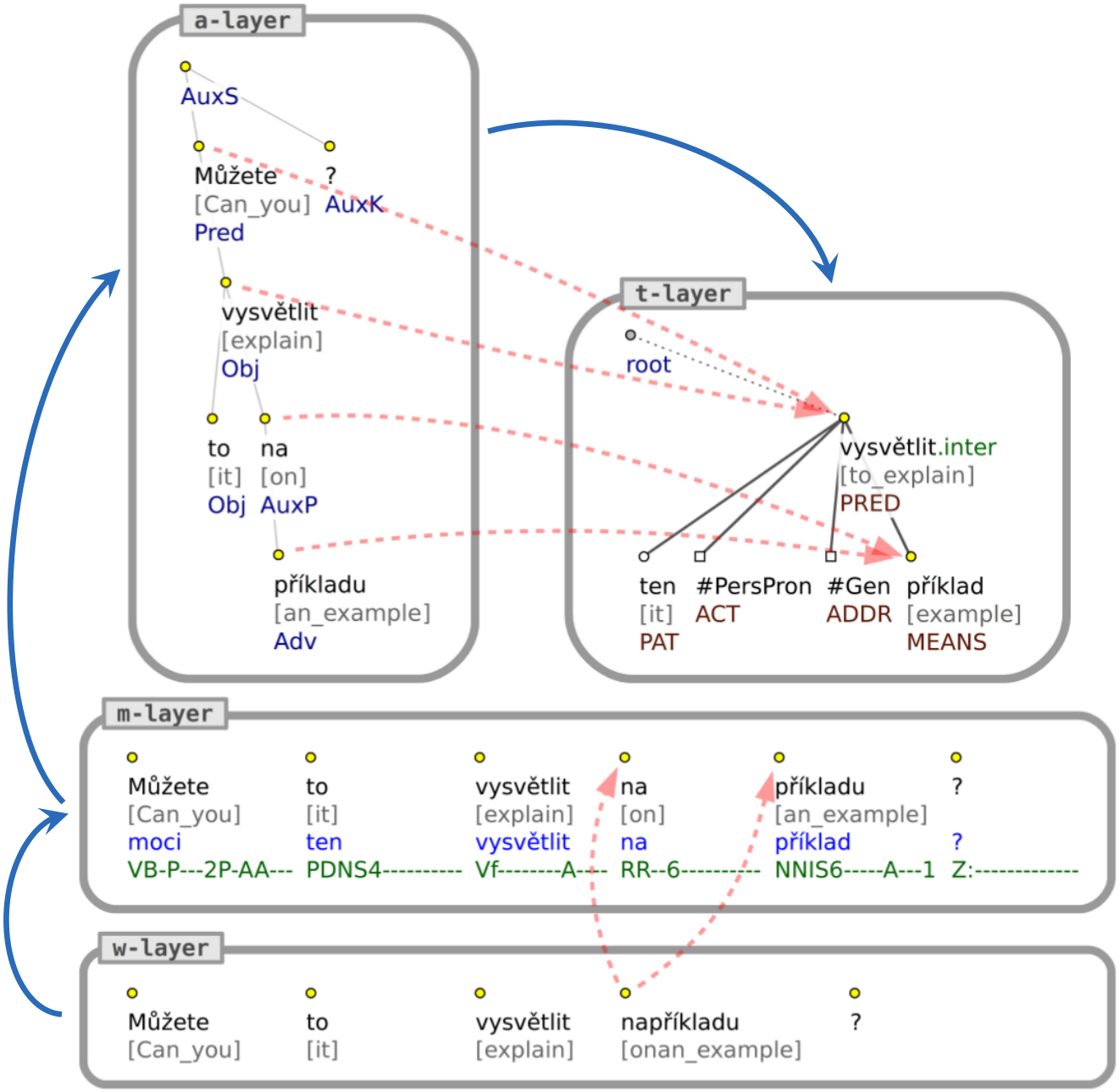
Every node of the tectogrammatical representation is assigned a complex label:

(i) lexical value of the word

(ii) the "morphological" grammatemes (values of morphological categories): Feminine, Plural with nouns, Preterite etc. with verbs etc.

(iii) the 'functor': Actor, Patient,..., kinds of Adjuncts; plus 'subfunctors' for a more detailed classification of functors

(iv) the topic-focus articulation attribute (TFA), three values for contextual boundness (contrastive and non-contrastive cb, nb)


a-layer
AuxS
Můžete
[Can_you]
Pred
?
AuxK
vysvětlit
[explain]
Obj
to
[it]
Obj
na
[on]
AuxP
příkladu
[an_example]
Adv
t-layer
root
vysvětlit.inter
[to_explain]
PRED
ten
[it]
PAT
#PersPron
ACT
#Gen
ADDR
příklad
[example]
MEANS
m-layer
Můžete [Can_you] moci VB-P---2P-AA---
to [it] ten PDNS4----------
vysvětlit [explain] vysvětlit Vf--------A---
na [on] na RR--6----------
příkladu [an_example] příklad NNIS6-----A---1
? ? Z:----------
w-layer
Můžete [Can_you]
to [it]
vysvětlit [explain]
napříkladu [onan_example]
?

ANNOTATION OF INFORMATION STRUCTURE (TOPIC-FOCUS ARTICULATION)

WHY?

Semantic relevance:

John only introduced Mary to SUE

(a) … only to SUE (i.e. to nobody else),

(b) only Mary to SUE (i.e. nobody else to nobody else),

(c) only introduced Mary to SUE (he did not do anything else)

Everyone in this room knows at least two languages.

At least two languages are known by everyone in this room.

Dogs must be CARRIED.

DOGS must be carried. = Carry DOGS.

➔ common denominator: position in deep order: TOPIC – FOCUS

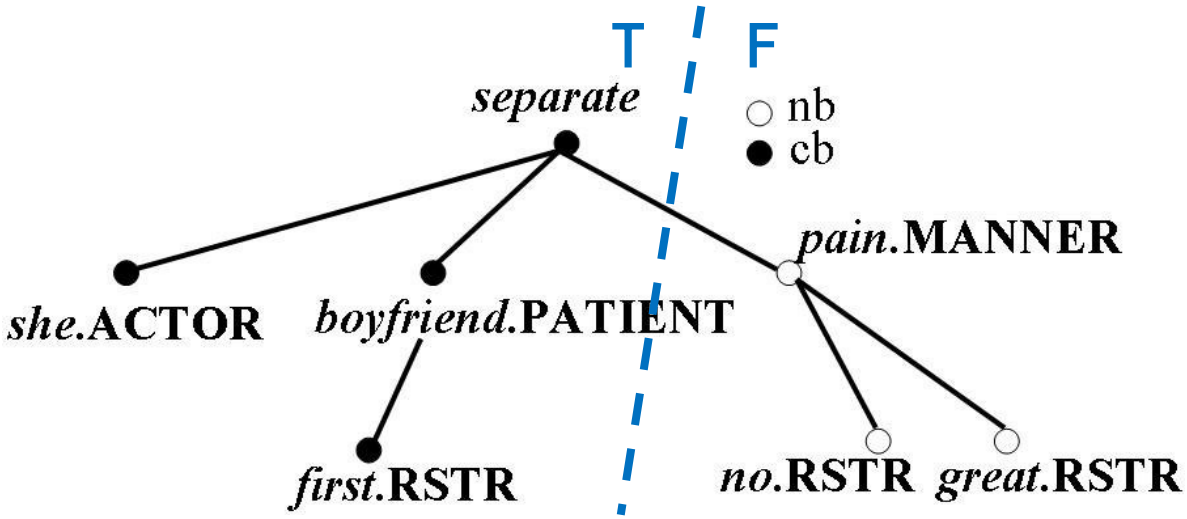

on the right
Hold the handrail Keep clear of the edges
Take extra care with children
Dogs must be carried
No smoking


DOGS MUST
BE CARRIED

Zpráva velmi příznivá – Češi udělali revoluci.
A méně příznivá – revoluci udělali Češi.
Excellent news: The Czechs made a revolution.
And the bad news: The revolution-Acc. made the Czechs-Nom

TOPIC AND FOCUS IN PDT

In FGD: boundary between Topic and Focus – algorithm, based on the TFA values cb and nb



(PRECEDING CONTEXT: *IRENA HAD TWO BOYFRIENDS AND SEPARATED FROM BOTH OF THEM.*)
SHE HAD SEPARATED FROM HER FIRST BOYFRIEND / WITH NO GREAT PAIN.

ANNOTATION OF COREFERENCE

- In the theoretical approach – textual coreference and discourse relations go beyond the tectogrammatical level (level of linguistic meaning)

… but:

- In PDT: both relations are annotated on the tectogrammatical trees → the information in the tree helps

- Grammatical coreference: reflexive pronouns, relative pronouns, subjects of verbs of control, 1st and 2nd pers. pronouns
- Textual coreference : specific vs. generic
- Asociative links: bridging anaphora:
  - Whole x part
  - Set x element(s) of a set
  - Object x function (*team – coach*)
  - Pragmatic contrast (*this year – last year*)
  - Specific relations: e.g. author – piece of work

Nedoluzhko (2013), Annotation of coreference and bridging, Zikánová et al. (2015), Discourse and Coherence

ANNOTATION OF DISCOURSE RELATIONS

Annotation of discourse relations – based on the PDTB approach

discourse connective: a predicate of a binary relation – it takes two text spans (mainly clauses or sentences) as its arguments and signals a semantic relation between them

(i) Explicit connectives
(ii) AltLex (alternative lexicalization)
(iii) Annotation of genres
(iv) Work in progress: implicit relations and so-called No-rel: No signal can be found

Prasad et al. (2008), The Penn Discourse Treebank; Poláková et al. (2014), Genres; Zikánová et al. (2015), Discourse and Coherence; Rysová et al. (2016), Prague Discourse Treebank

3. DEPENDENCY

3.1 RETROSPECTIVE

- continental **syntactic theories** versus the mainstream syntactic approaches on the other side of the Atlantic
- Bloomfield (1933): names of the main constituents of the sentence, NP and VP
- Chomskyan approach: originally: exclusively based on the concept of immediate constituents, later: the notion of head in the X-bar theory

  N, V, Adj and P(rep) as possible heads ➔ may be interesting unless the set of basic categories grows beyond some reasonable limit

BUT: gradual development of the X-bar theory ➔ practically any constituent may act as the head

TRADITIONAL CONSTITUENTS NOT SUITABLE: SPECIFICATION OF INFORMATION STRUCTURE (TFA)

John spent two weeks in Mallorca.

Test for T/F: questions

How did John spend his holidays last year? *Last year John spent / two weeks in Mallorca.*

What did John do last year? *Last year John / spent two weeks in Mallorca.*

Combinatorial Categorial Grammar (Mark Steedman) "floating constituents" or "non-standard" constituents

John / ate the beans
John ate / the beans
John / ate \ the beans

Hajičová (1972), Aktuální členění větné [Topic-Focus Articulation], Steedman (1991), Structure and Intonation

3.2 THE REPERTOIRE OF DEPENDENCY RELATIONS

- operational criteria – rather than intuitions
- economy - as **small number** of elementary units as possible
- valency structure : the core of predication that is more straightforwardly represented in a dependency rather than in a phrase-structure (constituency) approach
- closer to the **logical representation** (predication: n-ary logical predicate)

Arguments vs. Adjuncts:

Several tests proposed: omission, reformulation with a relative clause "which happened", substitution by "do so"

In **Functional Generative Description**:

Two "axis": (a) arguments vs. adjuncts

(b) semantically obligatory vs. optional

Panevová (1974), On Verbal Frames; Hajičová (1983), Remarks on the Meanings of Cases; Panevová (2003), Some Issues of Syntax and Semantics of Verbal Modifications; Panevová (2016), In Favour of the Argument-Adjunct Distinction

(I) ARGUMENTS VS. ADJUNCTS

(i) Arguments - Adjuncts

Criteria:

(a) can the given modification depend on **every** verb?

(b) can the given modification be **repeated** with a single governor?

modifications that satisfy neither (a) nor (b) = **arguments** (inner participants): Actor, Patient, Addressee, Origin, Effect

(special position of Actor)

modifications that satisfy both = **adjuncts**

some **ontological blocking** – e.g. “purpose” with verbs of the change of state:

(?) John fell ill in order to be punished for his sins.

SURFACE REALIZATION

arguments usually take a certain form in the surface form of the sentence influenced by the verb, the morphemic form is governed by the requirements of the verb (the so-called *rection* or *government*)

adjuncts: their form is usually not governed by the governing verb –

Děti přišly domů ... do školy ... na hřiště

the children came home ... to school ... at the playground

(II) (SEMANTICALLY) OBLIGATORY VS. OPTIONAL

(ii) Semantically obligatory vs. optional:

''Dialogue test'': absence in the surface is not decisive: present in the deep structure and (maybe) deleted on the way to the surface (Fillmore (1977) "complete scene")

John has already arrived.

Where to? *I do not know. – WHERE-to is obligatory

From where? I do not know. –WHERE-from is optional

Mother is knitting a sweater.

From what? I do not know. – ORIG is optional

Panevová (1974), On Verbal Frames; Fillmore (1977), The Case for Case Reopened

VALENCY FRAME

	obligatory	optional
Arguments	+	+
Adjuncts	+	-

Valency frame: a set of modifications classified as valency slots of the lexical item

(a) Every modification classified as an **argument**: enters the frame, either obligatory or optional (with a special mark):

obligatory: *to require* smth.PAT, *to believe* sbd.ADDR, *to remember* sbd.PAT

optional: *to buy* smth: PAT.obl (to sbd.ADDR.opt);

(b) Obligatory **adjuncts** in the frame:

the operational test is applied

the sentence is often grammatically incorrect without them: **John aimed*

ISSUE AT STAKE: GENERALIZED OBLIGATORY VALENCY MEMBERS

"Generalizations": the argument is obligatory but its lexical setting is generalized

= a group of persons/objects/circumstances typical/usual for the given position

This dog does not bite good people. – *This dog does not bite* (= anybody). -> PAT_{Gen}

Jane sells (sg-PRS-IPFV) *in the Bata store*

Sell ACT PAT_{Gen} $ADDR_{Gen}$

What does she sell? - goods typical for Bata store

To whom does she sell? - to typical customers of the store

➔ Patient and Addressee: obligatory arguments

Panevová (2001), Valency Frames; Panevová (2003), Some Issues of Syntax and Semantics

4. Grammar and Lexicon

The **connection** of grammar and lexicon:

- quite apparent already in the earlier works of C. Fillmore (1966; 1968): case grammar,

  explicitly follows up Tesnière's notion of valency

- the concept of valency: crucial, reflects the fundamental aspect of grammatical information in the lexicon:

  the valency frame = a part of the lexical entry,

  in which the obligatory and optional arguments and obligatory adjuncts of the given head word are registered

Tesnière (1959), Elements ...; Fillmore (1966), Towards a Modern Theory of Case; Fillmore (1968), The Case for Case

VALENCY DICTIONARY - VALLEX

VALLEX 3.0 (2016, first edition 2008):

4586 Czech verbs corresponding to 2722 lexemes (= 6711 lexical units)

aspect counterparts: if counted separately = 10 821 lexical units

Two components:

(i) data component – the valency behaviour of individual verbal lexemes

(ii) grammatical component: grammatical rules for possible changes in valency structure:

to derive a marked usage of the given unit = passive, reflexive, reciprocal, and to determine which valency slots correspond to the same situational participant

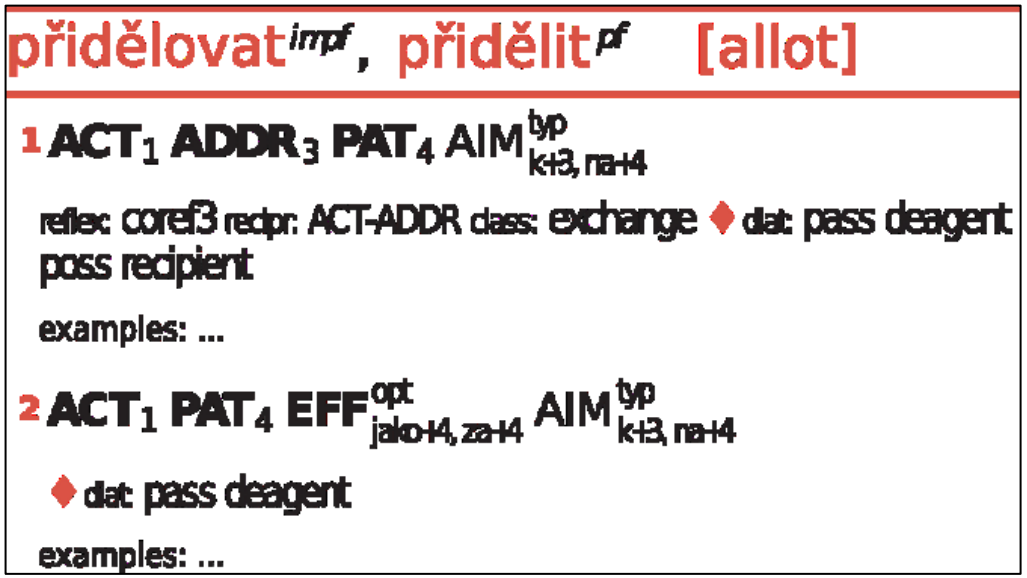
Apart from the valency frame: other syntactic and syntactico-semantic information: e.g. control, grammaticalized alternations (diathesis, reflexivity, reciprocity),

With every lexical unit: a special attribute indicating the possibility of application of an alternation

With selected lexical items: lexicalized alternations (e.g. conversion)

Lopatková et al. (2016), VALLEX; Lopatková, and Kettnerová (2016), Alternations

VALLEX (CONT.): EXAMPLE



- *přidělit* [allot]
- *Obecní úřad přidělil rodině byt*
  [The town council allotted a flat to the family]
- *Rodina-ADDR dostala od obecního úřadu-ACT přidělen byt-PAT*
  [Family-ADDR got from the town council-ACT allotted flat-PAT]

Lopatková et al. (2016), VALLEX

PDT-VALLEX

PDT-VALLEX: an on-line tool ➔ to check the existing valency frames and to add the new ones, for verbs, but also some adjectives and nouns

- originated "on the road"- during the annotation process, but systematically checked: "bottom–up, data-driven approach"
- linking of **each verb occurrence** in the PDT to the lexicon -> to verify the the valency lexicon entry against the corpus

7,127 verbs (11,933 valency frames), http://lindat.mff.cuni.cz/services/PDT-Vallex/

- headword (zřeknout se = abandon)
- valency frame

zřeknout se

zřeknout se$_{7x,7x}$ **ACT**(1) **PAT**(2)

(zřící se, odmítnout, vzdát se) • *zřeknout se sestavení vlády*

members & their labels (ACT, PAT), their obligatoriness feature and the surface form (1-nom., 2-gen.)

- an example: a fragment of a Czech sentence taken from the PDT
- Notes: used to delimit the meaning of the individual valency frames within the entry, also synonyms, antonyms and aspectual counterparts

Hajič, Panevová, Urešová, Bémová, Kolářová & Pajas Petr (2003). PDT-VALLEX: Creating a Large-coverage Valency Lexicon for Treebank; Hajič & Urešová (2003). Linguistic Annotation: from Links to Cross-Layer Lexicons; Urešová (2011). Valence sloves v Pražském závislostním korpusu [Valency of Verbs in PDT]

ENGVALLEX AND CZENGVALLEX

EngVallex: largely manual adaptation of the PropBank Lexicon, relabeling of the arguments, obligatoriness added, frames; links to PropBank lexicon wherever possible

size: 7,148 valency frames for 4,337 verbs

CzEngVallex:

- alignments between Czech and English valency frames, in a stand-off mode
- 20835 aligned valency frame pairs (verb senses) which are translations of each other, alignment of their arguments as well
- annotation done on bilingual data from the parallel **PCEDT corpus**.

Urešová, Fučíková, Šindlerová (2016), CzEngVallex: A Bilingual Czech-English Valency Lexicon, Cinková (2006), From PropBank to EngVallex: Adapting the PropBank Lexicon to the Valency Theory of the Functional Generative Description..

VALENCY OF NOUNS

* PLÁN 'PLAN'

: id: blu-n-plán-1
+ ACT(gen,poss;obl) PAT(gen,na+acc,o+loc,pro+acc,inf,aby,že,dcc; obl)
-derived: blu-v-plánovat-1
-gloss: **záměr, úmysl 'intention, aim'**
-example: plány vystavět.PAT podnik 'plans to construct a processing plant'
-control: ACT, ex
-class: mental action
-type: **abstract result**

: id: blu-n-plán-2
+ AUTH(gen,poss;typ) PAT(gen,poss; obl)
-derived: blu-v-plánovat-1
-gloss**: mapa, nákres 'map, layout'**
-example: *plán města*.PAT 'city plan'
-class: mental action
-type: **concrete**

5. CHALLENGES

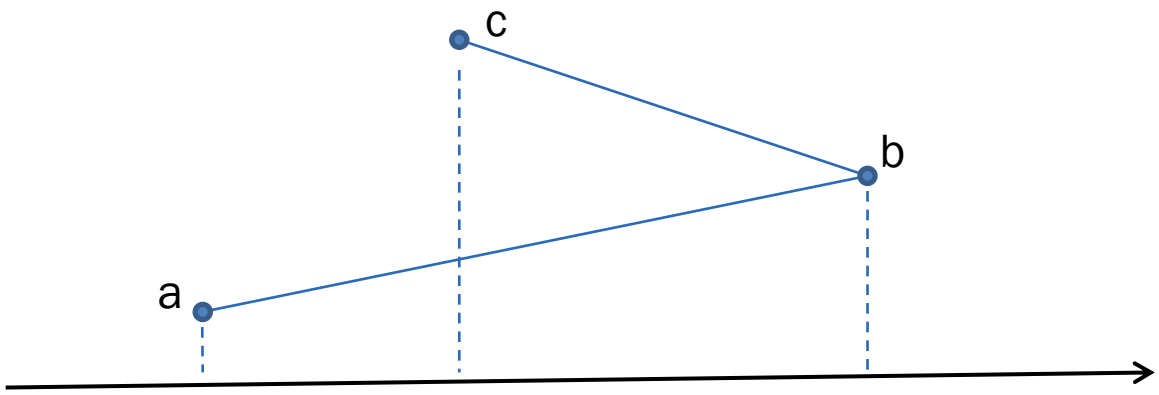
5.1 NON-PROJECTIVITY

the relation between **syntactic structure and word order** (the discontinuity of constituents) – a matter of discussions within all syntactic theories

unbounded dependencies

long-distance dependencies

Non-projective construction:



CONDITION OF PROJECTIVITY (CONT.)

- The **more** restricted the formal syntactic description is, the more valuable are the observations based on it
- BUT: there are many non-projective constructions **in the surface** shapes of the sentences

one possibility in a multilevel framework:

projectivities preserved on the surface level but a **strong** projectivity condition on the deep level

➔ to formulate *"movement rules"* specified as a transition from projective underlying trees to the surface representations in which the condition of projectivity cannot be applied

= a rather strong hypothesis that has to be verified and made more precise on the basis of a systematic empirical research

Sgall (1998), On the usefulness of Movement Rules

NON-PROJECTIVITIES IN THE PDT: SOME STATISTICS

On the analytical (surface) level:

73.088 sentences (with 1.255.590 occurrences of words incl. punctuation marks)

The condition of projectivity not met with 23.691 pairs of words (1,9 %)
at least one non-projective edge: in 16.920 sentences (23,2 % of all sentences)

three groups:

A. types of structures that can be specified on TR on the basis of **specific** grammatical properties, or limited lexical groups (37%)

B. divided syntagms, mostly due to the position of contrastive topic (6%)

C. auxiliary structures, mostly predicates with function words forms (57%) – not relevant for the tectogrammatical level

Zeman (2004), Neprojektivity v PZK [Non-Projectivities in PDT]; Hajičová et al. (2004), Issues of Non-Projectivity in PDT

5.2 Coordination and Apposition

High frequency of parsing errors : e.g. their impact on quality of dependency-based MT is substantial ➔ number of solutions in treebanks designs

Modifiers shared by all conjuncts

Mary came and cried

Nested coordination :

John and Mary or Sam and Lisa

Shared modifiers can be coordinated:

(big and cheap) (apples and oranges) – ambiguity!!! ((big and cheap) apples) and (oranges)

Intricate structures when combined with ellipsis

Careful semantic analysis necessary:

red and white wine x *red and white flag of Poland*

five dogs and cats x *five dogs and five cats*

COORDINATION (CONTD.)

Different models proposed:

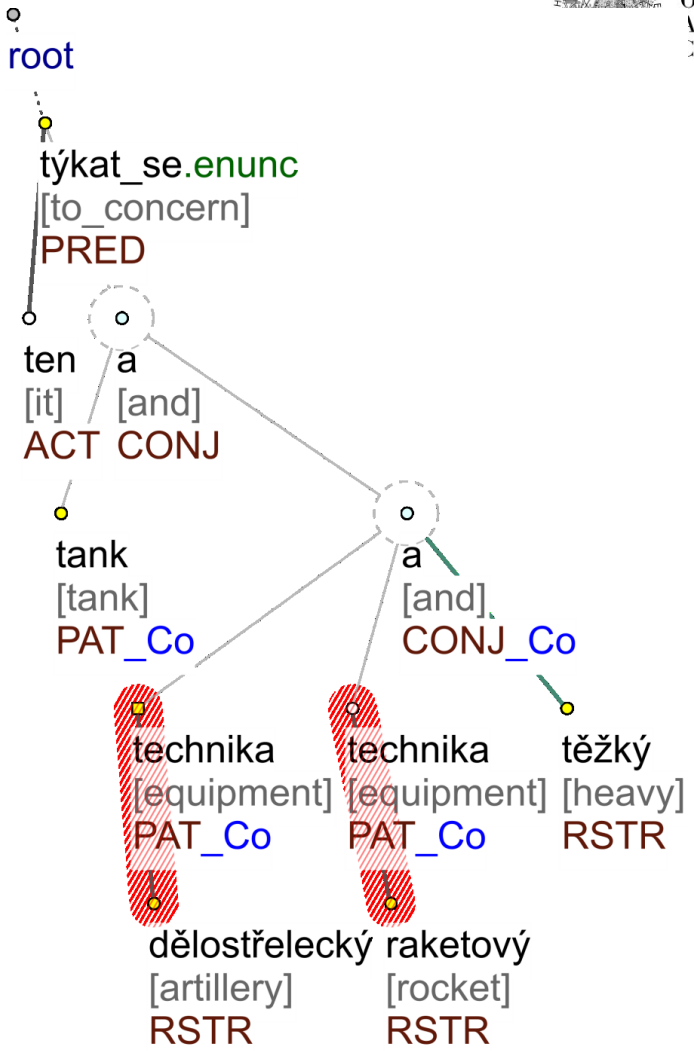
MTT (Mel´chuk) – the first conjunct is the head of the CS, the second attached as a dependent of the first one etc., coordinating conjunction attached under the penultimate conjunct, and the last conjunct is attached under the conjunction

FGD (Prague, Sgall) – coordination and apposition is not considered to be a dependency relation as they cannot be captured by the usual binary directional dependency relation =>

non-dependency, *"horizontal" structure*, possibly n-ary and/or nested, but still unidirectional

- all elements have (in the standard dependency sense) a common governor (the only exception: coordinated predicates which naturally have no common governor)
- the coordinated elements can also have common dependent(s)

In PDT: all conjuncts attached under the coordinating conjunction – along with shared modifiers which are distinguished by special attributes



Týkalo se to tanků a těžké dělostřelecké a raketové techniky.
[It concerned tanks and heavy artillery and rocket equipment.]

5.3 RECONSTRUCTION OF SURFACE DELETIONS

Ellipsis: an omission of a unit at the surface shape of the sentence but necessary for the semantic interpretation =>

= ellipsis as an empty place in a sentence that has not been occupied by a lexical unit

> Fillmore (2000) "understood but missing"
>
> Kayne (2005) silent elements, "elements that despite their lack of phonetic realization seem to have an important role in the syntax of all languages"

➔ a mismatch between syntax and semantics:

> most explicitly reflected in multilevel frameworks, e.g.
>
> Meaning-Text Theory (MTT), I. Mel'chuk
>
> Functional Generative Description (FGD), P. Sgall

Fillmore (2000), Silent anaphora; Kayne (2005), Movement and silence

EXAMPLES: "KROMĚ"(BESIDES)

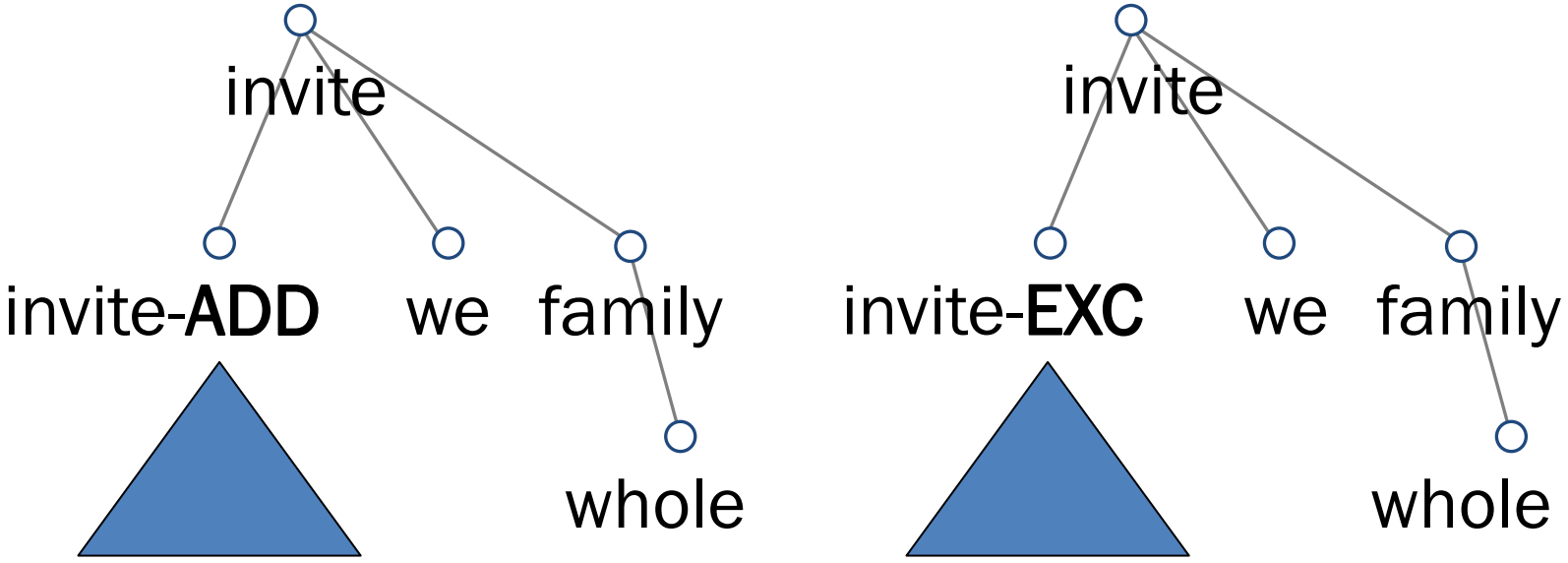
Czech: *kromě* [besides]

Kromě Jany pozveme celou rodinu.

Besides Jane we will invite the whole family.

(i) Jane will be invited (too) (Addition)

(ii) Jane will not be invited (Exclusion)

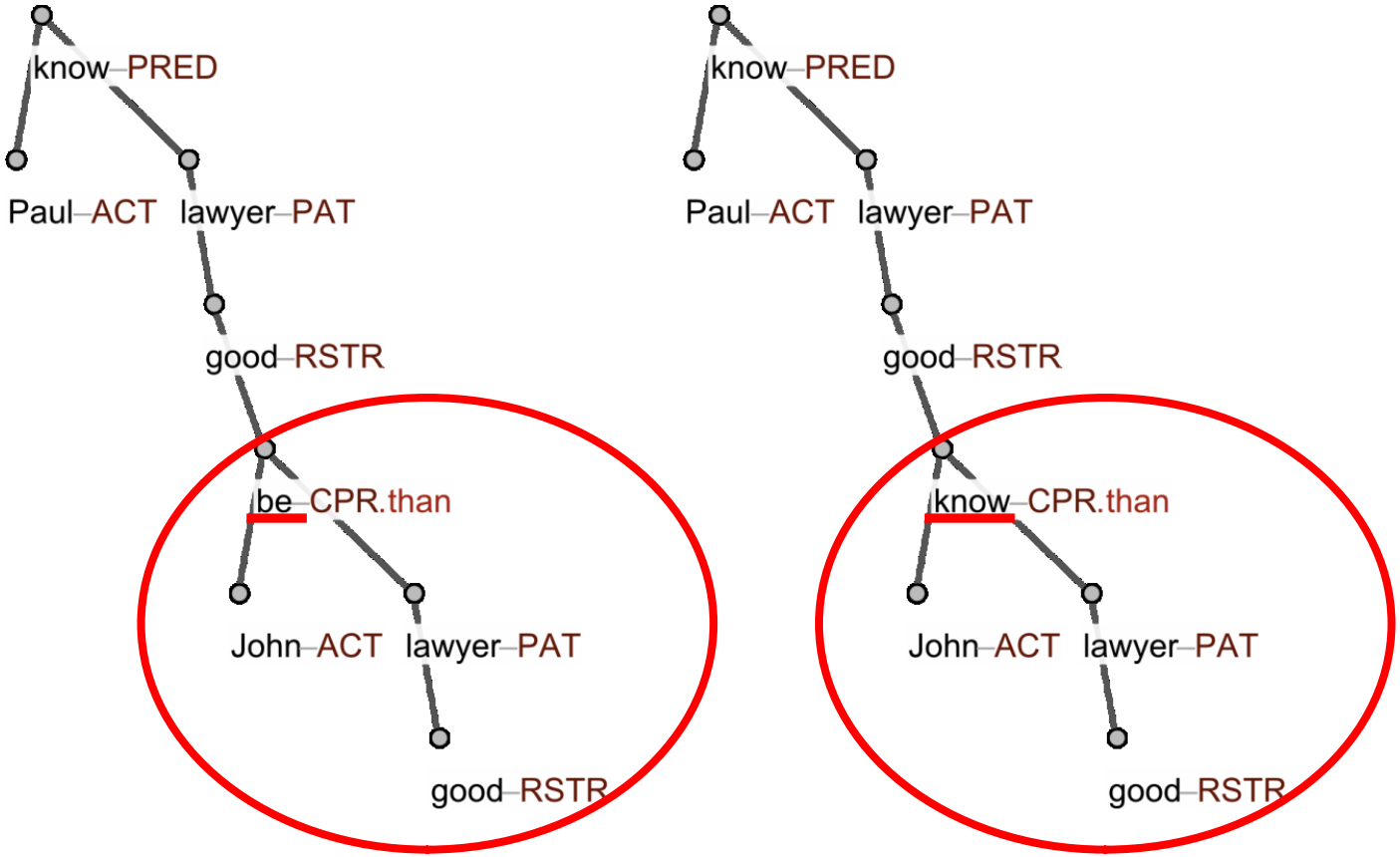


EXAMPLES: COMPARISON

Paul knows a better lawyer than John.

(a)... a better lawyer than John (is a lawyer)

(b)... than John knows (a lawyer)



ELLIPSIS IN PDT

on the analytical (surface) level: missing **governor**:

ExD (extra-dependency) to the node for which the governor is missing -> challenge for the study of deletions

on the tectogrammatical level: missing **dependents**

nodes are added to the tree, with special labels according to the type of ellipsis

Special technical solution for **coordination** structures

Hajičová, Mikulová and Panevová (2015), Reconstructions of Deletions; Hajič et al. (2015), Deletions and Node reconstructions

SUMMARY (1)

- Theoretical description of language: both “deep” and “dependency” are crucial
  - Deep:
    - Remove ambiguity
    - Deletions, information structure

    → multi-layered system of representation
  - Dependency:
    - “Economy”
    - Information structure
    - Non-standard constituents

SUMMARY (2)

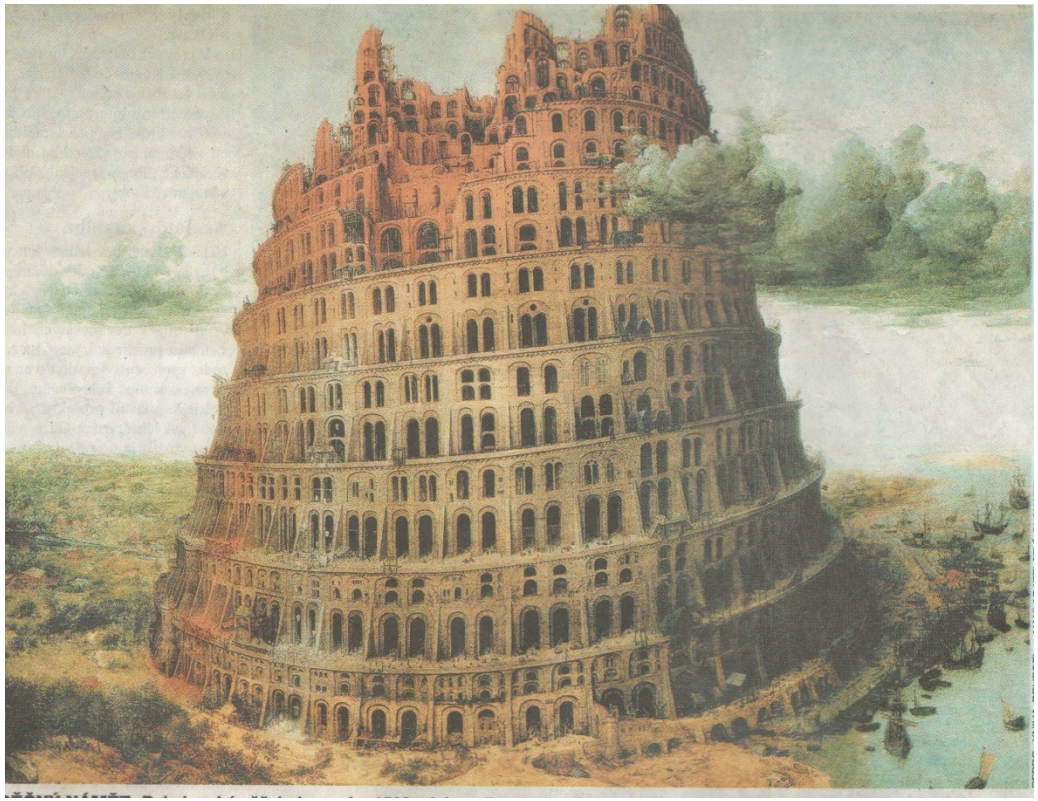
- From the point of NLP: despite undisputable recent progress in NLP which relies more on computational methods than linguistic representations or features, we believe that for true understanding having an adequate theory is worth the effort

SUMMARY

- Ars longa, vita brevis
- … occasio praeceps, experimentum periculosum, iudicium difficile
- „… art is long, life is short, opportunity fleeting, experiment dangerous, judgement difficult“

Hippocratés

THANK YOU FOR YOUR ATTENTION!



REFERENCES:

- Bejček E, Hajičová E, Hajič J, et al. (2013), Prague Dependency Treebank 3.0. Data/software, Univerzita Karlova v Praze, MFF, ÚFAL, Prague, Czech Republic, http://ufal.mff.cuni.cz/pdt3.0/
- Cinková S. (2006), From PropBank to EngVallex. Adapting the PropBank Lexicon to the Valency Theory of the Functional Generative Description. In: Proc. of LREC 2006, 2170-2175.
- Dokulil M. and F. Daneš (1958), K tzv. významové a mluvnické stavbě věty {On the semantic and and grammatical structure of the sentence], In: O vědeckém poznání soudobých jazyků, Prague, 231-246,.
- Fillmore C. J. (1966), Toward a modern theory of case. The Ohio State University Project on Linguistic Analysis, Rep. No. 13, 1-24.
- Fillmore C. J. (1968), The Case for Case. In: E. Bach and R. Harms (eds.), Universals in Linguistic Theory, New York, 1-88.
- Fillmore C. (1977), The case for case reopened. In K. Heger and J. Petöfi (eds.), Kasustheorie, Klassifikation und sematische Interpretation. Hamburg, Buske, 3-26.
- Fillmore, C.J. (1999). Silent Anaphora, Corpus, FrameNet and Missing Complements, Paper presented at the TELRI Workshop, November 1999, Bratislava
- Hajič J (1998) Building a Syntactically Annotated Corpus: The Prague Dependency Treebank. In Issues of Valency and Meaning. Studies in Honour of Jarmila Panevová, ed. Hajičová E, Karolinum, Prague, pp 106-132
- Hajič J., Hajičová E., Mírovský J. and J. Panevová (2016), Linguistically Annotated Corpus as an Invaluable Resource for Advancements in Lingutic Research: A Case Study, PBML 106, 69-124
- Hajič J., Hajičová E., Mikulová M. and , Mírovský J., Panevová J. and D. Zeman (2015), Deletions and node reconstructions in a dependency-based mutlilevel annotation scheme. In: *16th International Conference on Computational Linguistics and Intelligent Text Processing, Lecture Notes in Computer Science*, ISSN 0302-9743, 9041, pp. 17-31, Springer, Berlin / Heidelberg
- Hajič J., Hajičová E., Panevová J. et al.(2011), Prague Czech-English Dependency Treebank 2.0
- Hajič J., Panevová J., Urešová Z., Bémová A., Kolářová V. and P. Pajas Petr (2003), PDT-VALLEX: Creating a Large-coverage Valency Lexicon for Treebank Annotation. In: Proceedings of The Second Workshop on Treebanks and Linguistic Theories, Vaxjo University Press, Vaxjo, Sweden, pp. 57-68.
- Hajič J. and Z. Urešová (2003). Linguistic Annotation: from Links to Cross-Layer Lexicons. In: Proceedings of The Second Workshop on Treebanks and Linguistic Theories, Vaxjo University Press, Vaxjo, Sweden, pp. 69-80.
- Hajičová E. (1972), Aktuální členění větné a nejnovější vývoj transformační gramatiky [Functional Sentence Perspective and the latest developments in transformational grammar], Slovo a slovesnost XXXIII, 229-239.

REFERENCES:

- Hajičová E. (1983), Remarks on the Meanings of Cases Prague Studies in Mathematical Linguistics 8, 149-157
- Hajičová E.(2012), What we have learned from complex annotation of Topic-Focus Articulation in a large Czech corpus. Echo des études romanes, Vol. VIII/1, 51-64
- Hajičová E., Havelka J., Sgall P, Veselá K. and D. Zeman (2004): Issues of Projectivity in the Prague Dependency Treebank. In: *The Prague Bulletin of Mathematical Linguistics*, pp. 5-22
- Hajičová E., Mikulová M. and J. Panevová (2015), Reconstructions of Deletions in a Dependency-Based Description of Czech: Selected Issues. In Proc. of the Third Int. Conference on Dependency Linguistics, Uppsala, 131-140.
- Hajičová E., Partee B. H., Sgall P. (1998), Topic-Focus Articulation, Tripartite Structures, and Semantic Content, Kluwer Academic Publishers, Dordrecht
- Halliday M.A.K. (1967), Notes on transitivity and theme in English. Journal of Linguistics,3, 37-81, 199-244; 4(1968), 179-215.
- Kayne R. S. (2005), *Movement and Silence*, Oxford University Press
- Kettnerová, V., Lopatková, M.: At the Lexicon-Grammar Interface: The Case of Complex Predicates in the Functional Generative Description. In Depling 2015: Proceedings of the Third International Conference on Dependency Linguistics, Uppsala University, Uppsala, Sweden, pp. 191-200. 2015.
  https://ufal.mff.cuni.cz/%7Elopatkova/2016/docs/15-4-DepLing-LVC.pdf

  Kettnerová Václava, Lopatková Markéta, Bejček Eduard: The Syntax-Semantics Interface of Czech Verbs in the Valency Lexicon. In Proceedings of the 15th EURALEX International Congress, Department of Linguistics and Scandinavian Studies, University of Oslo, Oslo, Norway, pp. 434-443. 2012 https://ufal.mff.cuni.cz/%7Elopatkova/2014/docs/4-12-euralex-sep.pdf
- Klímová Jana, Kolářová Veronika, Vernerová Anna (2016): Towards a Corpus-based Valency Lexicon of Czech Nouns. In: GLOBALEX 2016: Lexicographic Resources for Human Language Technology. GLOBALEX workshop 2016, pp. 1-7.
- Kolářová, V. (2014). Special valency behavior of Czech deverbal nouns. In O. Spevak (Ed.) Noun Valency, Amsterdam: John Benjamins, 19-60.
- Lopatková M. et al. (2016), Valenční slovník českých sloves VALLEX [Valency Dictionary of Czech verbs]. Prague: Karolinum
- Lopatková Markéta, Kettnerová Václava: Alternations: From Lexicon to Grammar And Back Again. In Proceedings of the Workshop on Grammar and Lexicon: Interactions and Interfaces (GramLex), The COLING 2016 Organizing Committee, Ōsaka, Japan, pp. 18-27, 2016 http://ufal.mff.cuni.cz/~lopatkova/2017/docs/16-CoLing-sep.pdf
- Lopatková M., Plátek M. and P.,Sgall (2007), Towards a Formal Model for Functional Generative Description: Analysis by Reduction and Restarting Automata. PBML 87, pp. 7-26.

REFERENCES:

- Nedoluzhko A.: Generic noun phrases and annotation of coreference and bridging relations in the Prague Dependency Treebank. In: 51st Annual Meeting of the Association for Computational Linguistics Proceedings of the 7th Linguistic Annotation Workshop & Interoperability with Discourse, Omnipress, Inc, Sofija, Bulgaria, pp. 103-111, 2013
- Panevová J. (1974), On Verbal Frames in Functional Generative Description , PBML 22, 3-40.
- Panevová J. (1980), Formy a funkce ve stavbě české věty [Forms and Functions in the Structure of Czech sentence], Prague
- Panevová J. (2001), Valency Frames: Extension and Re-examination. In: Studies in Syntax and Semantics of Slavonic Languages. Papers in Honour of Andrzej Boguslawski on the occasion of his 70th birthday (eds. V.S.Chrakovskij et al.), Studia Slavica Oldenburgensia 9, Oldenburg, 325-340.
- Panevová J. (2003), Some Issues of Syntax and Semantics of Verbal Modifications. In: Proceedings MTT 2003, Paris: Ecole Normale Supérieure, 2002, 139-146.
- Panevová J. (2016), In Favour of the Argument-Adjunct Distinction (from the perspective of FGD), PBML 106,21-30
- Panevová J. et. al. (2014), Mluvnice současné češtiny 2. Syntax češtiny na základě anotovaného korpusu.[Syntax of Czech on the basis of an annotated corpus]. Prague, Karolinum.
- Poláková L., Jínová P. and J. Mírovský (2014): Genres in the Prague Discourse Treebank. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14),* Reykjavik, Iceland, 1320-1326
- Poláková L., Mírovský J., Nedoluzhko A., Jínová P. , Zikánová Š. and E. Hajičová (2013): Introducing the Prague Discourse Treebank 1.0. In: *Proceedings of the 6th International Joint Conference on Natural Language Processing*, Asian Federation of Natural Language Processing, Nagoya, Japan, 91-99
- Popel M., Mareček D., Štěpánek J., Zeman D. and , Z.Žabokrtský (2013): Coordination Structures in Dependency Treebanks. In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pp. 517-527, Association for Computational Linguistics, Sofija, Bulgaria,
- Prasad R., Dinesh N., Lee A. et al. (2008): The Penn Discourse Treebank 2.0. In: *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, Morocco, 2961–2968
- Rysová, M., Synková, P., Mírovský, J., Hajičová, E., Nedoluzhko, A., Oceľák, R., Pergler, J., Poláková, L., Scheller, V., Zdeňková, J., Zikánová, Š (2016).: *Prague Discourse Treebank 2.0.* ÚFAL MFF UK, Prague, Czech Republic, Lindat/Clarin: http://hdl.handle.net/11234/1-1905, Dec 2016

REFERENCES:

- Sgall P. (1977), Meaning of Sign, Cognitive Content and Pragmatics. Journal of Pragmatics 1, 269-282..
- Sgall P. (1980) Case and Meaning, Journal of Pragmatics 4, 525-536.
- Sgall P.(1998), On the usefulness of movement rules, In: Proc. of the 16th Int. Congress Of Lingustists, OAris. Oxford Elsevier Sciences.
- Sgall P. (2006), Multifarious Aspects of Language. Karolinum, Prague.
- Sgall P. and E. Hajičová (1970), A Functional Generative Description: Background and Framework. PBML 14, 3-38.
- Sgall P., Hajičová E. and J. Panevová (1986), The Meaning of the Sentence in Its Semantic and Pragmatic Aspects, Reidel, Dordrecht
- Steedman M. (1991), Structure and Intonation. Language 67, 260-296.
- Tesnière L (1959), Elements de syntaxe structurale. Paris, Klinkesieck.Urešová Zdeňka (2011a). Valence sloves v Pražském závislostním korpusu [Valency of Verbs in the Prague Dependency Treebank], UFAL, Prague. 229pp.,
- Urešová Zdeňka (2011b). Valenční slovník Pražského závislostního korpusu [Valency dictionary PDT-Vallex], UFAL, Prague. 375pp.
- Urešová Z., Fučíková E. and J. Šindlerová (2016). CzEngVallex: a Bilingual Czech-English Valency Lexicon, The PBML 105, pp. 17-50.
- Urešová Zdeňka, Fučíková Eva, Hajič Jan and Jana Šindlerová (2015), CzEngVallex. URL http://hdl/handle.net/11234/1-1512. LINDAT/CLARIN digital library at UFAL, Charles University, Prague.
- Zeman D. (2004), *Neprojektivity v Pražském závislostním korpusu (PDT)* (technical report). ÚFAL/CKL MFF UK (http://ufal.mff.cuni.cz/techrep/tr22.pdf)
- Zikanova Š. et al. (2015), Discourse and Coherence, UFAL MFF UK, Prague

- Abbreviations:
- PBML: The Prague Bulletin of Mathematical Linguistics
- UFAL MFF UK: The Institute of Formal and Applied Linguisticis, Faculty of Mathematics and Physics, Charles University, Prague