

Delimiting Adverbial Meanings

A corpus-based comparative study on Czech spatial prepositions
and their English equivalents

Marie Mikulová, Veronika Kolářová, Jarmila Panevová, Eva Hajičová
UFAL, Charles University, Prague, Czech Republic



- Introduction
- Research in delimiting adverbial meanings
 - Language-oriented Functional Generative Description
 - Main principle: substitutability of forms
- A comparative study
 - Prague Czech-English Dependency Treebank
 - Czech spatial prepositions and their English equivalents
- Conclusions

Functional Generative Description and Prague Dependency Treebanks



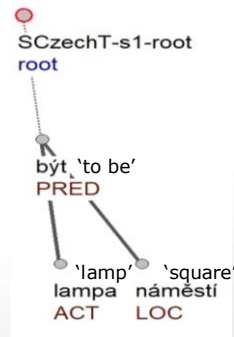
• Functional Generative Description

- A language-oriented rather than ontology-oriented dependency-syntax approach
 - an exclusive focus on the way how the given language in its structure reflects the reality
- A detailed corpus-based research

• Prague Dependency Treebanks

- Written Czech (PDT 3.0), spoken Czech (PDTSC)
- Prague Czech-English Dependency Treebank (PCEDT)
- Underlying dependency structure of the sentences
 - content words together with their auxiliary words, such as prepositions, are represented by nodes of a treeshaped graph
 - **functors**: types of dependency relations (functions) attached to all nodes

Lampa je na náměstí.
'The lamp is on the square.'

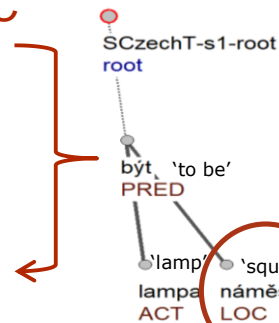


- General adverbial meaning: “where” - functor LOC

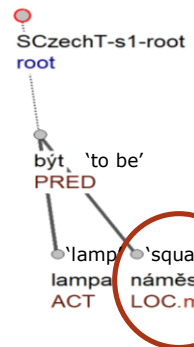
- (1) *Lampa je na náměstí.* ‘The lamp is *on the square.*’
- (2) *Lampa je za náměstím.* ‘The lamp is *behind the square.*’
- (3) *Lampa je mimo náměstí.* ‘The lamp is *out of the square.*’
- (4) *Lampa je před náměstím.* ‘The lamp is *in front of the square.*’
- (5) *Lampa je blízko náměstí.* ‘The lamp is *near the square.*’

...

- Subtle meanings: “on”, “behind”, “near”, etc.
- We need **subfunctors!**



PDT representation
Now: only LOC functor



Our goal:
Functor: LOC
Subfunctor: mid

- (6) *Lampa je uprostřed náměstí.* ‘The lamp is *in the middle of the square.*’



- Distinguishing between **cognitive content** and **linguistic meaning**



- (1)
- U domu* roste strom.
‘There is a tree **by the house**.’
- Vedle domu* roste strom.
‘There is a tree **next to the house**.’
- Blízko domu* roste strom.
‘There is a tree **near the house**.’

one content - different meanings

- (2)
- Na chalupě* máme bazén.
‘We have got a pool **at the cottage**.’



different contents - one meaning

- No one-to-one relation** between the functor-subfunctor combination and its formal expression (e.g., a prepositional group)
 - one syntactic function can be expressed by several different forms whereas one form can be used to express different syntactic functions



- (1) *Na chalupě máme bazén.*
 'We have got a pool **at the cottage.**'
- (2) ~~*V chalupě máme bazén.*~~
 'We have got a pool **in the cottage.**'



- (1) *Na chalupě máme bazén.*
 'We have got a pool **at the cottage.**'
- (2) *V chalupě máme bazén.*
 'We have got a pool **in the cottage.**'

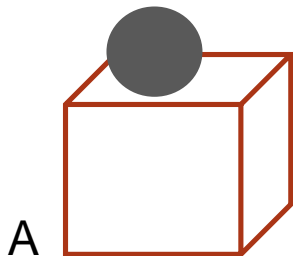


- (1) *Na chalupě máme bazén.*
 'We have got a pool **at the cottage.**'
- (2) ~~*V chalupě máme bazén.*~~
 'We have got a pool **in the cottage.**'

- **Examination of substitutability** of forms **na** 'in/on/at' and **v** 'in/on'
 - the form **na** 'in/on/at' can be substituted by the preposition **v** 'in' only in case the real localization is "inside" (B)
 - impossible to determine without the knowledge of the situation or without a clue from the context
 - the meaning "inside" does not appertain to the form **na** 'in/on/at' ("at the given place")
 - the form **na** 'in/on/at' introduces an object as a whole, covering several different localizations

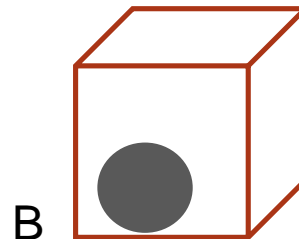
Testing synonymy

Main principle: **Substitutability of forms**



(1) *Míč je na krabici.*
'The ball is **on the box.**'

(2) ~~*Míč je v krabici.*~~
'The ball is **in the box.**'



(1) ~~*Míč je na krabici.*~~
'The ball is **on the box.**'

(2) *Míč je v krabici.*
'The ball is **in the box.**'

- Examination of substitutability of forms **na** 'in/on/at' and **v** 'in/on' (another context)
 - forms **v** 'in' and **na** 'in/on/at' are not substitutable if the given place is a **3D object**
 - form **na** 'in/on/at': "on the surface" (A)
 - form **v** 'in': "inside" (B)

Subfunctors for localizations “within the given place” (LOC functor)



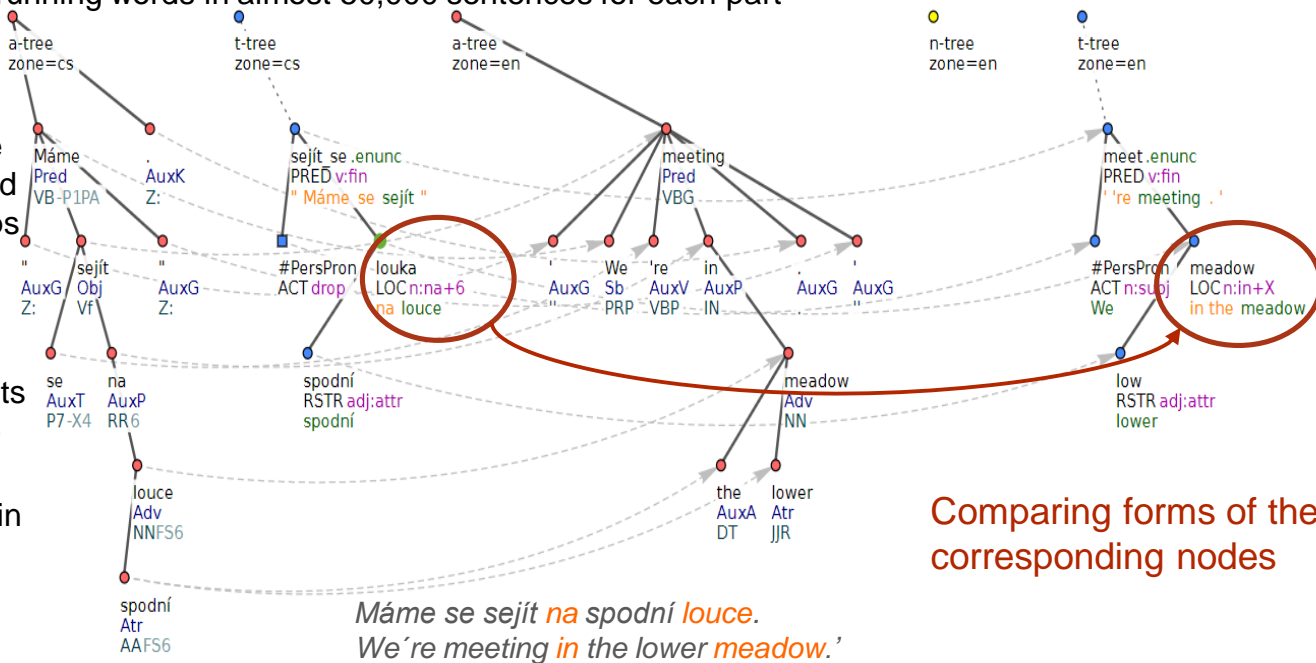
Function		Form	Example
functor	subfunctor		
LOC “where”	<i>inside</i>	<i>v</i> ‘in’	<i>V chalupě</i> máme bazén. ‘We have a pool in the cottage .’
		<i>uvnitř</i> ‘inside’	<i>Uvnitř</i> těch <i>zemí</i> jsme navštívili známá města. ‘ Inside the countries we visited the famous cities.’
	<i>surface</i>	<i>na</i> ‘in/on/at’	<i>Míč je na krabici</i> . ‘The ball is on the box .’
	<i>at the given place</i>	<i>na</i> ‘in/on/at’	<i>Na chalupě</i> máme bazén. ‘We have got a pool at the cottage .’
		<i>u</i> ‘by/at’	<i>Přespal jsem u kamaráda</i> . ‘I slept over at (my) friend’s (place).’

Prague Czech-English Dependency Treebank

- English: the Wall Street Journal section of the Penn Treebank, Czech: translated
- 1.2 million running words in almost 50,000 sentences for each part

Steps:

- Cz-adverbials with the LOC functor expressed by prepositional groups
- Their most frequent En-equivalents
- Cz-En pairs sorted ~ forms of En-equivalents
- Manual assignment of the subfunctors to the respective adverbials in each Cz-sentence



Comparing forms of the corresponding nodes

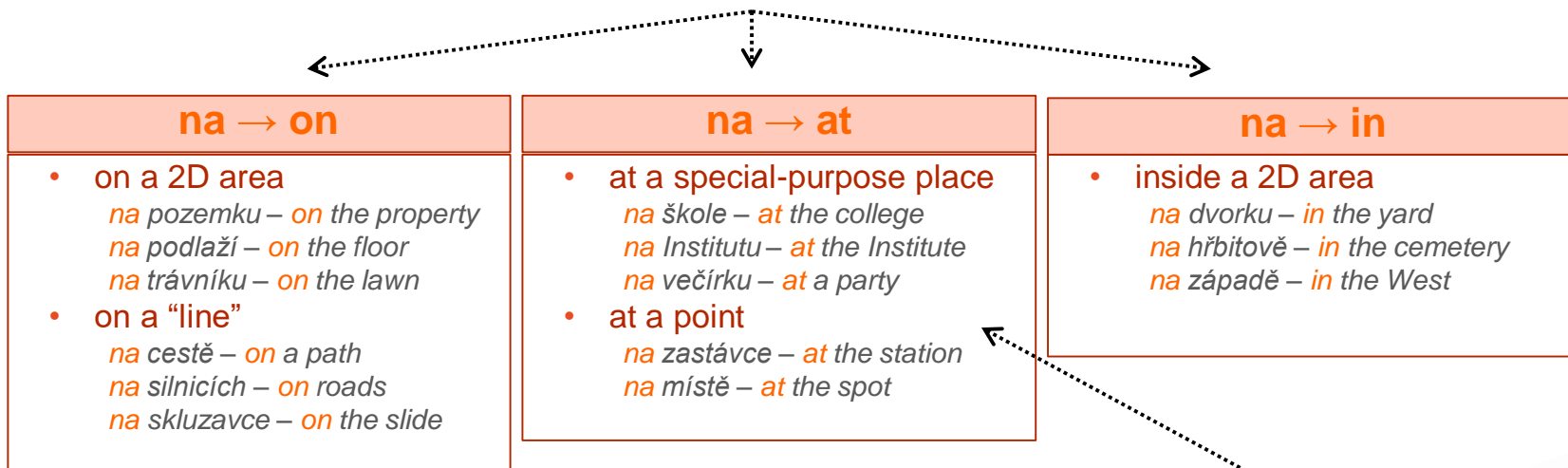
A comparative study: Czech spatial prepositions and their English equivalents



CZ-form	Subfunctor	EN-form	Number of pairs	Example
na 331	surface	on	4	na stole – on the desk
		at	1	na moři – at the sea
	at the given place	on	147	na trávníku – on the lawn
		in	93	na světě – in the world
		at	86	na škole – at a school
v 3061	inside	in	2913	ve věži – in the tower
		at	88	v továrně – at a factory
		on	60	v televizi – on television
u 18	at the given place	at	8	u agentury – at the agency
		in	6	u soudu – in the court
		with	3	u příbuzných – with relatives
		on	1	u soudu – on the court

Equivalents for *na* with the subfunctor “at the given place”

- The subfunctor “at the given place” covers several localizations
 - not fixed in Czech, in contrast to English
 - a **vague character** of the Czech preposition *na*



The English preposition *at* makes it possible to differentiate further semantic nuances

High degree of oscillation between forms and meanings



- In both languages
 - **inside a 2D area versus on the surface of a 2D area**
 - *v / na ulicích – on / in the streets*
 - *v / na světě – in the world*
 - *na ostrově – on / in the island*
 - **inside a 3D object versus at a special-purpose place**
 - *v / na škole – in / at the school*
 - *na trhu - in / on / at the market (1)-(3)*

(1) *There is finally some sort of sense **in the market**.*

(2) *It had to buy sugar **on the world market** to meet export commitments.*

(3) *They graze **at the Farmers Market**, a combination food court and grocery store.*

Conclusion



- Our analysis has confirmed that **both languages structure the reality in a different way**
 - English spatial prepositions can express more specific features of reality than the Czech ones
 - two Czech prepositions **na / v** vs. three English ones (**at / in / on**)
- More precise definitions of subfunctors for Czech
- The deep syntactic representation of PDTs decreases the “distance” between languages, yet there does not exist a universal set of subfunctors.



We thank all our sponsors.



Thank you for your attention.

