

Delimiting Adverbial Meanings. A corpus-based comparative study on Czech spatial prepositions and their English equivalents

Marie Mikulová

Faculty of Mathematics and Physics
Charles University, Prague
mikulova@ufal.mff.cuni.cz

Veronika Kolářová

Faculty of Mathematics and Physics
Charles University, Prague
kolarova@ufal.mff.cuni.cz

Jarmila Panevová

Faculty of Mathematics and Physics
Charles University, Prague
panevova@ufal.mff.cuni.cz

Eva Hajičová

Faculty of Mathematics and Physics
Charles University, Prague
hajicova@ufal.mff.cuni.cz

Abstract

The data of the Prague Czech-English Dependency Treebank (a member of the family of Prague Dependency Treebanks) have served as a basis for the comparative study of delimiting adverbial meanings of the local relation “within the given place”. The Czech prepositional groups containing the prepositions *v*, *na*, and *u* with the above meaning are compared with their English equivalents, using a more subtle differentiation into three semantic subgroups of “inside”, “on the surface” and “at the given place”. Our analysis confirms that though every language structures the reality in a different way, certain tendencies may be observed in the relation of the forms and their functions that eventually result in a more detailed classification. The contribution presents results of an ongoing work.

1 Introduction

The description of adverbial meanings has a long tradition in linguistics, varying in its attention to detail (e.g. Quirk et al., 1985; for Czech: Šmilauer, 1947). However, it is well known that the traditional classification of adverbials is not fine-grained enough, either in theoretical description or for NLP tasks.

In the multi-layered scenario of Prague Dependency Treebanks (Sect. 2), linguistic meaning is captured by the deep syntactic layer, where the syntactic relations are represented by the so-called functors. However, the functors capture relatively general categories. E.g., all the following adverbials *na stole* ‘on the table’, *pod stolem* ‘under the table’, *za stolem* ‘behind the table’, *poblíž stolu* ‘near the table’, etc. are represented by a single functor with a static meaning “where” (functor LOC). It is obvious that a differentiation among the partial meanings (“on the given place”, “under the given place”, “behind the given place”, etc.) is needed for a more precise representation of the sentence meaning and for its translation to another language. In order to describe these fine-grained distinctions, a set of so-called subfunctors has been considered (Mikulová et al., 2017).

The area of spatial meanings is wide. It includes the general meanings of “where”, “which way”, “to where” and “from where” (which we capture by functors), and also their subtle meanings (“inside”, “on the surface”, “next to”, “under”, etc.), for which we propose subfunctors.¹ In the paper, we analyze only a narrow, highly problematic set of meanings within the LOC functor (“where”). We focus on the specification of spatial adverbial meanings expressed by prepositional groups (Sect. 3). Our Czech-English parallel data (Sect. 2) make it possible to compare corresponding expressions in the two languages and to explore differences in forms and meanings, in particular those expressing localization “within the given place” (Sect. 4). We believe that such an analysis will help us to evaluate the universality and language specificity of the suggested subset of adverbial meanings and thus to make the description of this subset for Czech more precise.

¹For the delimitation of the functors, the lexical meaning of the verb and its valency properties may be a useful clue, whereas subfunctors are primarily expressed within prepositions.

2 Theoretical Background and Data Resources

2.1 Functional Generative Description

We base our investigation on the theoretical framework of the Functional Generative Description (Sgall et al., 1986), a language-oriented rather than ontology-oriented dependency syntax theory. As for the relationship between language meaning and ontological content, the FGD works with language meaning in the sense of structural linguistics, treating meaning as a linguistically structured phenomenon. When describing attributes necessary for the layer of language meaning, we inevitably tackle the boundary between meaning and content, for example by differentiating homonymy (properties of a form in relation to meaning) and vagueness (properties of meaning in relation to content). We search for testable criteria to be able to account for these distinctions and also to specify synonymy (Sect. 3).

Compared with other descriptions of spatial relations,² our approach is characterized especially by the following aspects:

- An exclusive focus on the way how the given language in its structure reflects the reality
- Dependency syntax approach
- A detailed corpus-based research.

2.2 Prague Czech-English Dependency Treebank

The ideas of the Functional Generative Description were applied in the annotation scenario of the Prague Dependency Treebanks (Hajič et al., 2017).³ The Prague treebanks are complex linguistically motivated corpora with interlinked hierarchical layers of standoff annotation (on morphological, surface and deep syntactic layer). The pilot Prague Dependency Treebank (Hajič et al., 2018) was built in 1996 through 2018. A slightly modified scenario was then used for the annotation of the other treebanks.

The Prague Czech-English Dependency Treebank (PCEDT; Hajič et al., 2012), which is used for our comparative study, is an annotated Czech-English corpus. The English part consists of the Wall Street Journal section of the Penn Treebank (Marcus et al., 1993). Czech part was translated from the English source sentence by sentence.

3 Methodology of Delimiting Adverbial Meanings

Our analysis of adverbial meanings is based on the assumption that there is no one-to-one relation between the underlying syntactic function represented by the functor-subfunctor combination and its formal expression, in this case the preposition(al group). One syntactic function can be expressed by several different forms whereas one form can be used to express different syntactic functions.

Analyzing fine-grained adverbial meanings, we apply the following principles 3.1, 3.2, and 3.3.

3.1 Substitutability of forms

When deciding which forms are synonymous and thus can be described by the same subfunctor we test whether the forms are substitutable in different contexts and how their meaning is influenced by the substitution. The forms may be semantically the same, they can only partially correspond to each other, or they are not substitutable at all.

E.g., the two prepositions *v* ‘in’ and *na* ‘in/on’ are substitutable in cases when the semantic distinctions between them are obscured due to the fact that they form a prepositional group with words denoting objects that do not distinguish „inside“ and „surface“, cf. (1), or this distinction is excluded by a broader sentential context, cf. (2). However, when it comes to localizations beyond the meaning “inside”, the preposition *na* ‘in/on’ cannot be substituted by *v* ‘in’ (cf. (3), where the greenhouse is supposed to be placed in the garden).

- (1) Umíte se dobře zorientovat **v mapě**? /→*na mapě*
'Are you able to read the map (lit. to orient yourself in a map)? /→*on a map*'

²There is a brief list of various analysis of the spatial prepositions: Bennett, 1975; Herskovits, 1986; Aurnague, 1995; Garrod et al., 1999; Lindstromberg, 2010; Talmy, 2006; Vandeloise, 1991; etc.

³<https://ufal.mff.cuni.cz/pdt-c>

- (2) Přespali jsme tam **na té chatě**. /→v chatě /≠u chaty
'We slept at the cottage. /→in the cottage /≠by the cottage.'
- (3) Mám **na chatě** skleník. /?v chatě
'I have got a greenhouse at my cottage. /?in the cottage'
- (4) Má **na hlavě** čepici. /*v hlavě
'He has a cap on his head. /*in his head'

The preposition *na* 'in/on' can be substituted by the preposition *v* 'in' only in case the real localization is „inside“, which is however impossible to determine without the knowledge of the situation or without a clue from the context. We conclude from this that the semantic feature „inside“ (subfunctor *inside*) does not appertain to the preposition *na* 'in/on'. The preposition *na* 'in/on' introduces an object as a whole, covering several different localizations within the given place (subfunctor *at the given place*). Similarly, forms *v* 'in' and *na* 'in/on' are not substitutable if the given place is a 3D object⁴ and the real localization is “surface”, cf. (4); thus the meaning “on the surface” (subfunctor *surface*) only appertains to the preposition *na* 'in/on'.

3.2 Partial synonymy

When delimiting subfunctors, we differentiate forms that are typical for the given function from those that are untypical for this (e.g. secondary prepositions). The untypical forms are always associated with certain connotations which do not arise with the typical ones. The test of substitutability is thus directed from the untypical forms to typical ones, e.g. *uvnitř* 'inside' → *v* 'in', cf. (5). A substitution in the opposite direction thus does not work in general, cf. (6).

- (5) **Uvnitř těchto zemí** jsme navštívili hlavní a známá města. /= v těchto zemích
'*Inside these countries* we visited the capitals and some famous cities. /= in these countries'
- (6) Byli jste někdy **v zahraničí**? /*uvnitř zahraničí
'Have you ever been abroad (**in a foreign country**)? /*inside a foreign country'

3.3 Disjunction of forms

One function (subfunctor) can be expressed by two or more forms that are not substitutable (in which case their meaning has to be inferable from the context). However, more forms do not imply more subfunctors. This is the case of forms *u* 'by' and *na* 'in/on' used for localization “within the given place”; cf. (2) and (7). The form *u* 'by' can only be used in this meaning in contexts restricted to a certain group of lexical items, bearing some animate and institutional features.

- (7) Přespal jsem **u kamaráda**. /*na kamarádovi/*v kamarádovi
'I slept over **at (my) friend's (place)** / *on (my) friend / *in (my) friend.'

4 Comparative Study: Czech Spatial Prepositions and Their English Equivalents

In any language, prepositions for expressing localization are few in number but allow for a wide range of uses; this discrepancy presents a challenge for semantic analysis of spatial prepositions in a cross-linguistic perspective (Levinson - Wilkins, 2006).⁵ Based on the material of the PCEDT corpus (Sect. 2.2) we compare formal realizations of the corresponding deep syntactic units, focusing on the most frequently analyzed area of spatial meanings, namely localization “within the given place”. Applying

⁴Whenever we refer here to a 2D or 3D object, we have not in mind the real dimension, but we refer rather to the speaker's actual conception of the given place.

⁵Studies exploring the way how Czech and English structure spatial relationships focus especially on equivalents of particular prepositions in a parallel corpus data, either from the Czech-English perspective (Novotná, 2010; the preposition *na* 'on/in'), or from the English-Czech perspective (Kirschner, 1974; the preposition *in*). Investigating English equivalents of the most frequent Czech prepositions (i.e. *na* 'on/in', *v* 'in' and *s/se* 'with'), a semantic analysis is also carried out by Klégr et al. (2012), who classify spatial meanings 'where' into (i) location on the surface, (ii) a point in the space and (iii) a point inside the space. A systematic contrastive analysis of the meaning of English forms vis-à-vis their Czech counterparts is given by Strnadová in Dušková et al. (2006); she observes that English spatial prepositions can express more specific features of reality than the Czech ones. This corresponds to the observations of Hruška (1976), who states that English spatial adverbials display ability to differentiate more precisely various notions of place by means of a wider choice of prepositions (cf. *between* and *among*). The relations between Czech and English forms are also described in Czech textbooks of English (cf. e.g. Vít, 2019).

the principles described in Sect. 3, we have subcategorized this localization into a set of three subfunctors associated with the corresponding Czech forms, as illustrated in Table 1.

In the Czech part of the PCEDT corpus we have searched for adverbials with the LOC functor (depending on a verb) expressed by prepositional groups containing the prepositions *v* ‘in’, *na* ‘in/on’ or *u* ‘by’,⁶ and then looked for their most frequent equivalents in the English part. The Czech-English pairs of sentences were then sorted out according to the form of the English equivalent. Finally, we have manually assigned the subfunctor of the local specification to the respective adverbials in each Czech sentence (see Table 2).⁷

Subfunctor	Form	Example
<i>inside</i>	<i>v</i> ‘in’	(1), (6)
	<i>uvnitř</i> ‘inside’	(5)
<i>surface</i>	<i>na</i> ‘in/on’	(4)
<i>at the given place</i>	<i>na</i> ‘in/on’	(2), (3)
	<i>u</i> ‘by’	(7)

Table 1: Subfunctors for localizations “within the given place” (of LOC functor).

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

Table 2: Czech prepositions for localization “within the given place” and their English equivalents.

In spite of the fact that the collected material is not large, certain tendencies can be followed:

(A) The equivalent for *v* with the subfunctor *inside* is mostly the form *in* (e.g. **inside a 3D object**: *ve vozidlech – in cars*, *v garáži – in a garage*, *v košíku – in a basket*; **inside a 2D area**: *v regionu – in the district*, *v zemi – in the country*, *v Číně – in China*; **in a piece of art**: *v knize – in a book*, *ve filmu – in a film*, *v dopisech – in the letters*; **in a domain**: *v průmyslu – in the industry*, *v technologii – in technology*).

(B) The equivalent for *na* with the subfunctor *surface* is mostly the form *on* (e.g. **on the surface of a 3D object**: *na stole – on the table*, *na čepicích – on caps*, *na kopci – on a hill*).⁹

⁶We have not examined here the secondary preposition *uvnitř* ‘inside’.

⁷We exclude cases where the equivalent in the English sentence is not a prepositional group. Since the texts in the corpus are mostly mono-thematic (economic and political texts from journals), the lexically identical pairs are counted as a single case (e.g. in Table 2, 128 occurrences of *na trhu – in the market* are counted as a single case of the equivalence). We also exclude cases of annotation mistakes and we do not work with idiomatic and fixed phraseological expressions.

⁸For the most frequent occurrence of *v-in* (2913 pairs) the first 200 different pairs have been analyzed, other figures in the Table 2 are the total numbers of the given pairs in the material analyzed.

⁹There are only few examples in our data, but the observation is confirmed by the conclusions in Klégr et al. (2012).

Other English equivalents for the subfunctors *inside* and *surface* are rather rare (cf. Table 2) and concern an oscillation described below. Only two rather conspicuous subgroups expressing localization *inside* can be distinguished, both with the English form *on* corresponding to the Czech form *v*, i.e. **means of communication** (e.g. *ve vysílání – on a broadcast, v rádiu – on the radio, v televizi – on television*) and **transport** (e.g. *ve vlaku – on the train*).

(C) The equivalents for *na* with the subfunctor *at the given place* are almost evenly distributed among the forms *at, in, on*. The prevailing tendencies are as follows:

(C-i) The form *na* with the subfunctor *at the given place* is equivalent to *on* first of all with the localization **on a 2D area** (e.g. *na pozemku – on the property, na podlaží – on the floor, na trávníku – on the lawn*) and **on a “line”** (e.g. *na cestě – on a path, na silnicích – on roads, na skluzavce – on the slide*).

(C-ii) The form *na* with the subfunctor *at the given place* is equivalent to *at* in case the localization is understood as **a special-purpose place** (such as an institution or an event: *na škole – at the college, na Institutu – at the Institute, na večírku – at a party, na konferenci – at the conference*) and in case the location is understood as **a point** (e.g. *na zastávce – at the station, v centru – at the Center*).

(C-iii) The form *na* with the subfunctor *at the given place* is equivalent to *in* first of all in case of the localization **inside a 2D area** (e.g. *na dvorku – in the yard, na hřbitově – in the cemetery, na severozápadě – in the Northwest*).

(D) The equivalent forms for *u* with the subfunctor *at the given place* are the prepositions *at, in* and *with*. If the given location is **an institution**, all the above three forms may occur (e.g. *u agentury – at the agency, u soudu – in the court, u firmy – with the firm*). If the given location is **a person**, the equivalent is primarily the preposition *with* (e.g. *u příbuzných – with relatives, u ředitele – with the director*).

4.1 Discussion

The tendencies (A) and (B), i.e. a clear equivalence of the forms *v – in* and *na – on*, are very strong and support the differentiation of the opposite locations *inside – surface*. Originally, we have delimited the subfunctor *surface* as an opposition to the meaning of *inside* just with 3D objects (cf. Sect. 3). However, the tendency in (C-i) indicates a possibility to expand the scope of this subfunctor to localization “on the surface” of both 2D areas and 3D objects.

The tendency (C) confirms the vague character of the preposition *na* in Czech; it is evident that the subfunctor *at the given place* covers several meanings, which are not fixed in Czech, in contrast to English. The preposition *at* makes it possible to differentiate further semantic nuances in English, described here in a simplified way as localization at a special-purpose place or at a point, cf. (C-ii) and parallel Czech-English examples (8) and (9). In Czech, for the localization perceived as “at a special-purpose place” the preposition *u* (primarily expressing the localization „beside“) is used; however, its coverage is narrower than with the English form *at*, cf. (D).

- (8) Až dosud se inzeráty společnosti objevovaly téměř výlučně **v novinách** a časopisech.
Until now, the corporate ads have appeared almost exclusively **in newspapers** and magazines.
- (9) Podle podmínek smlouvy, která byla uzavřena **v novinách** Toronto Star, se 500 zaměstnanců...
Under the terms of the contract reached **at** Torstar **newspaper**, the 500 workers...

The analysis of our material has also demonstrated that both languages provide a high degree of contextual substitutability of two or even more forms expressing localization with a very slight difference in meaning (cf. the three English equivalents of the only Czech expression *na trhu* ‘in the market’ (10)-(12)). A localization can be perceived and structured in language in the different ways with different (language) meanings. Our material reflects a specially high degree of oscillation between the expression of the meanings **inside a 2D area** versus **on the surface of a 2D area** (e.g. *v ulicích/na ulicích – on the streets/in the streets, ve světě/na světě – in the world; na ostrově – on the island/in the island*) and **inside**

a 3D object versus **at a special-purpose place** (e.g. *ve škole/na škole – in the school/at the school, ve/u společnosti – in the company/at the company*).

- (10) There is finally some sort of sense **in the market**.
- (11) It had to buy sugar **on the world market** to meet export commitments.
- (12) They graze **at the Farmers Market**, a combination gourmet food court and grocery store.

There is also an appreciable established correlation between the given lexical unit and a certain preposition, which is especially frequent with toponyms but occurs also in other cases and which affects the validity of general tendencies (e.g. *na Havaji* (**v* Havaji) – *in Hawaii*).

5 Conclusion

Our analysis has confirmed that every language may structure the reality in a different way and that there may be an „*overwhelming diversity, and apparently endless mismatches between any two languages in both the formal coding of distinctions, and semantical basis for them*“ (Levinson – Wilkins, 2006, 550). Our analysis has also supported the conclusions of previously published studies that English spatial prepositions can express more specific features of reality than the Czech ones. The deep syntactic representation of the Prague Dependency Treebanks decreases the “distance” between languages, yet there does not exist a universal set of subfunctors. Cross-language studies help to explore the differences in structuring the reality and their description is useful for teaching and translation applications.

Acknowledgements

The research has been supported by the Czech Science Foundation under the project GA17-12624S and by the LINDAT/CLARIAH-CZ project of Ministry of Education, Youth and Sports of the Czech Republic (LM2018101). This research also has been using language resources and tools developed and/or stored and/or distributed by the LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (project LM2015071).

Reference

- Michel Aurnague. 1995. Orientation in French spatial expressions: formal representations and inferences. *Journal of Semantics*, 12(3), 239-268.
- David C. Bennett. 1975. *Spatial and Temporal Use of English Prepositions: An Essay in Stratificational Semantics*. Longman, London.
- Libuše Dušková et al. 2006. *Mluvnice současné angličtiny na pozadí češtiny*. Academia, Prague.
- Simon Garrod, Gillian Ferrier and Siobhan Campbell. 1999. In and on: investigating the functional geometry of spatial prepositions. *Cognition*, 72, 167-189.
- Jan Hajič, Eva Hajičová, Marie Mikulová and Jiří Mírovský. 2017. Prague Dependency Treebank. In: *Handbook on Linguistic Annotation*. Springer, Verlag, Berlin, 555-594.
- Jan Hajič, Eduard Bejček, Allevtina Bémová, Eva Buráňová, Eva Hajičová, Jiří Havelka, Petr Homola, Jiří Kárník, Václava Kettnerová, Natalia Klyueva, Veronika Kolářová, Lucie Kučová, Markéta Lopatková, Marie Mikulová, Jiří Mírovský, Anja Nedoluzhko, Petr Pajas, Jarmila Panevová, Lucie Poláková, Magdaléna Rysová, Petr Sgall, Johanka Spoustová, Pavel Straňák, Pavlína Synková, Maga Ševčíková, Jan Štěpánek, Zdeňka Uřešová, Barbora Vidová Hladká, Daniel Zeman, Šárka Zikánová, Zdeněk Žabokrtský. 2018. *Prague Dependency Treebank 3.5*. Institute of Formal and Applied Linguistics, LINDAT/CLARIN, Charles University, LINDAT/CLARIN PID: <http://hdl.handle.net/11234/1-2621>.
- Jan Hajič, Eva Hajičová, Jarmila Panevová, Petr Sgall, Ondřej Bojar, Silvie Cinková, Eva Fučíková, Marie Mikulová, Petr Pajas, Jan Popelka, Jiří Semecký, Jana Šindlerová, Jan Štěpánek, Jiří Toman, Zdeňka Uřešová and Zdeněk Žabokrtský. 2012. Announcing Prague Czech-English Dependency Treebank 2.0. In: *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*. European Language Resources Association, Istanbul, 3153-3160.

- Annette Herskovits. 1986. *Language and Spatial Cognition: An Interdisciplinary Study of the Prepositions in English*. Studies in Natural Language Processing, Cambridge: Cambridge University Press.
- Jiří Hruška. 1976. An Attempt at Linguistic Characterology of Prepositions in Present Day English in Comparison with Czech. *Brno Studies in English*, 12, 125–144.
- Zdeněk Kirschner (published anonymously). 1974. Some Problems of the Automatic Analysis of English Prepositional Constructions. In: *Automatické zpracování textů* (Natural Language Processing). SNTL, Prague, 86–156.
- Aleš Klégr, Markéta Malá and Pavlína Šedová. 2012. *Anglické ekvivalenty nejjfrekventovanějších českých předložek*. Karolinum, Prague.
- Seth Lindstromberg. 2010. *English Prepositions Explained*. John Benjamins, Amsterdam/Philadelphia.
- Stephen C. Levinson and David P. Wilkins. 2006. *Grammar of Space. Explorations in Cognitive Diversity*. Cambridge University Press, Cambridge.
- Mitchell Marcus, Beatrice Santorini and Mary A. Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank.
- Marie Mikulová, Eduard Bejček, Veronika Kolářová and Jarmila Panevová. 2017. Subcategorization of adverbial meanings based on corpus data. *Jazykovedný časopis*, 68(2), 268–277.
- Renata Novotná. 2010. The Czech preposition *na* and its English Equivalents. In: F. Čermák, K. Kučera, V. Petkevič (eds). *Intercorp: Exploring a Multilingual Corpus*. Lidové noviny, Prague, 138–145.
- Randolph Quirk, Sidney Greenbaum, Geoffrey N. Leech and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Longman, London.
- Petr Sgall, Eva Hajičová and Jarmila Panevová. 1986. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. D. Reidel Publishing Company, Dordrecht.
- Vladimír Šmilauer. 1947. *Novočeská skladba*. Ing. Mikuta, Prague.
- Leonard Talmy. 2006. The fundamental system of spatial schemas in language. In: B. Hampe (ed.). *From Perception to Meaning: Image Schemas in Cognitive Linguistics*. Mouton de Gruyter, Berlin, 37–47.
- Claude Vandeloise. 1991. *Spatial Prepositions: A Case Study from French*. The University of Chicago Press, Chicago/London.
- Marek Vít. 2019. *Help for English*. URL: <https://www.helpforenglish.cz/article/2006060502-mistni-predlozky-in-on-at>.