Investigating valency-changing prefixes in Czech and German using large syntactically annotated data

Hana Hledíková
Institute of Formal and Applied Linguistics
Charles University, Prague
hana.hledikova@ufal.mff.cuni.cz

1 Introduction

Valency-changing morphology is a phenomenon described for many typologically diverse languages (Haspelmath and Müller-Bardey, 2004; Haspelmath and Sims, 2010; Wunderlich, 2015). Morphological operations such as the addition of a derivational affix can have an effect on the number and type of arguments that a verb has, i.e., its valency / argument structure. This submission quantifies this phenomenon as it is manifested in verbal prefixation in Czech and German - two languages in which prefixation is the most productive type of verb-formation and in which, although there are individual differences in the behavior of the prefixes (cf. the fact that German prefixes can be both separable and unseparable), the addition of a prefix to a verb often has an effect on its valency (cf. examples 1, 2).

(1) auf etwas achten   > etwas   be-achten
     on sth.ACC see   > sth.ACC PREF-see
     'pay attention to'   'consider'

(2) věřit něčemu   > pro-věřit   něco
     trust sth.DAT   > PREF-trust sth.ACC
     'trust'   'check, make sure'

The relationship between prefixation and valency / argument strucutre has been investigated in both Germanic and Slavic languages (e.g., Stiebels, 1996; Wunderlich, 1997; Aldinger, 2004; Ramchand, 2004; Romanova, 2006). German separable prefixes are sometimes called particles, due to the fact that they can occur separately from the verb and are therefore analogical to verb–particle constructions such as those found in English (cf. German *aufgeben* 'give up' > *Ich gebe auf* 'I give up', and English *give up > I give up*). In this paper, we consider both separable and unseparable prefixes together. It has been noted that along with having an effect on telicity and perfectivity, prefixes and verb particles can alter the valency of the base verb, leading to the addition or deletion of an object, changes in the kind of entity expressed in the object, and other types of formal and semantic alternations (e.g., Zeller, 2001; McIntyre, 2003; Svenonius, 2004; Ramchand, 2008; Dehé, 2015; Toivonen, 2020). The patterns of alternations have been found to be similar in Slavic and Germanic languages, leading to proposals of a common underlying structural description for both (Svenonius, 2004; Ramchand, 2008).

However, the phenomenon is difficult to analyze using large data. Available valency dictionaries are limited in size, tend to include high frequency items and only have limited coverage of prefixed verbs (cf. Lopatková et al., 2016 for Czech, Schumacher et al., 2004 and Ziem et al., 2019 for German).

2 Method

To bypass this problem, we use an approach based on large syntactically parsed corpora. We used data from the CoNLL 2017 Shared Task (Ginter et al., 2017, 16GB of web data in each language) parsed using UDPipe (Straka and Straková, 2017)[1] to extract the formal characteristics and dependency relation tags of the nominal and clausal dependents of each verbal lemma, along with the frequency with which they occur.

Next, we extracted pairs made up of a prefixed verb and its corresponding unprefixed

---

[1]The labelled attachment score (LAS) reported in the shared task's results is 69.52.

verb	% with accusative object	pref. vs. unpref.
chodit	14.58%	
do-chodit	58.14%	43.56%
cenit	22.23%	
do-cenit	56.30%	34.07%

Table 1: Percentage of occurrences with the accusative object for *chodit - dochodit*; *cenit - docenit*.
pref. vs. unpref. = difference in the percentage of occurrences between the unprefixed and prefixed verb in the pair

dep. rel.	form	pref. vs. unpref.
obj	accusative	5.78%
obl	*v* + locative	-2.55%
obl	*k* + dative	2.20%
obl	*do* + genitive	1.18%
obj	genitive	1.10%

Table 2: Change in the occurrence of unprefixed vs. prefixed verbs with a given syntactic dependent for the prefix *do-*.[2]
the dep. rel. = dependency relation tag; obj = object; obl = oblique nominal

verb from a list of verbal lemmas annotated for their morphemic structure using large lexical resources (Žabokrtský, 2022; Kyjánek et al., 2021) combined with manual post-checking.

For each of the verb pairs, the difference in the percentage of occurrences with each dependent for the unprefixed verb and the prefixed verb was calculated (cf. the occurrence of an accusative object for two example pairs in Table 1). These differences were then used to calculate the average effect that each prefix has on the number of occurrences of each type of dependent (cf. example for the Czech prefix *do-* in Table 2).

3 Results

The analysis of the prominent differences reveals some general patterns across both languages. In both Czech and German, there are dependents whose frequency increases due to

---

[2]Positive values mean that the prefixed verbs have *n*-% more occurrences with the dependent. Negative percentage means that the prefixed verbs have *n*-% less occurrences with the dependent.

the locative meaning of the prefix – for instance, the locative meaning of Czech *vy-* 'out' and German *aus-* 'out' is associated with the presence of a prepositional phrase including the preposition *z / aus* 'from'. A prominent pattern that emerged from the quatitative analysis is that for almost all the prefixes, the prefixed verbs appear more frequently with a direct object in the accusative (cf. Tables 3, 4). This result, along with the fact that the prefixed verbs are telic and have the perfective aspect in Czech (where grammatical aspect is a property obligatorily encoded on the verb), is in accordance with approaches that describe prefixed verbs and particle verbs as denoting a resultative event with the resulting state being predicated over an affected entity expressed in the direct object (Zeller, 2001; McIntyre, 2003; Svenonius, 2004; Ramchand, 2008; Dehé, 2015). In some cases, the prefix adds an affected entity and resulting state to an event that does not previously include it, as in examples 3, 4.

(3) wohnen   >   etw   ab-wohnen
    live   >   sth.ACC PREF-live
    'live'   'wear sth out by living'

(4) tancovat   >   u-tancovat   někoho
    dance   >   PREF-dance sb.ACC
    'dance'   'dance sb (to exhaustion)'

Aside from the affected entity, the object may also express other concepts connected to the resultative event, such as a resultant entity (e.g., *singen* 'sing' > *etw er-singen* 'gain sth by singing', *stávkovat* 'strike' > *něco vy-stávkovat* 'gain sth by striking') or quantity (of time/resources/etc.) (e.g., *quatschen* 'chatter' > *Zeit ver-quatschen* 'waste time by chatter', *pracovat* 'strike' > *od-pracovat osm hodin* 'to work off eight hours'). The unprefixed verb may already express an activity directed at an affected entity, but the prefix adds a higher degree of affectedness along with introducing formal changes - the entity may be expressed by a prepositional phrase in the unprefixed verb, while the prefixed verb requires a direct object (e.g., *auf etw treten* 'step on sth.' > *etw zer-treten* 'destroy sth by steping on it', *řvát na*

prefix	% increase in proportion of obj:Acc	# verb pairs with increase in obj.Acc
an	10.36%	393
er	9.98%	199
ver	6.64%	518
be	6.15%	327
auf	5.72%	308
ab	5.67%	341
ent	5.66%	141
aus	5.45%	396
um	5.41%	242
ein	5.22%	286
über	3.97%	198
zu	3.75%	252
zer	3.75%	72
ge	1.43%	24
nach	1.42%	172
vor	0.70%	282
mit	0.69%	288
bei	0.48%	26
unter	0.33%	51

Table 3: German prefixes which lead to an increase in occurrence with an accusative object.

prefix	% increase in proportion of obj:Acc	# verb pairs with increase in obj.Acc
pod	18.35%	67
ob	17.45%	71
nad	12.53%	15
od	11.11%	269
roz	11.04%	282
pře	9.86%	260
u	9.07%	273
vy	8.92%	640
s	8.48%	206
pro	8.45%	250
o	8.42%	226
za	7.63%	442
na	6.33%	357
do	5.78%	209
po	5.45%	328
za	2.73%	436
vz	2.02%	20
při	0.89%	138

Table 4: Czech prefixes which lead to an increase in occurrence with an accusative object.

někoho 'shout at sb' > *se-řvat někoho* 'scold sb by shouting'). Some verbs of movement in which the landmark is expressed using a prepositional phrase in the unprefixed verb express the landmark as a direct object in the prefixed verb (and the prefix is typically cognate with the preposition used with the unprefixed verb), e.g., *um etw fliegen* 'fly around sth' > *etw um-fliegen* 'around-fly sth', *plavat pod něčím* 'swim under sth' > *pod-plavat něco* 'under-swim sth'. Along with some other types of structures that were found in the data for both languages, the high incidence of these analogical patterns where the degree of occurrence with the accusative object is higher for the prefixed verb supports the claims that have been made about the similarity of the effect of Slavic and Germanic prefixes/particles, namely that they are both able to add an object unselected by the base verb and lead to essentially resultative constructions (e.g., Svenonius, 2004; Ramchand, 2008).

4 Conclusion

This approach to investigating changes in verbal valency is advantageous in that it allows the use of large, automatically annotated data. The formal patterns which surface as the most important in the quantitative analysis can then subsequently be analyzed for their semantics. An analysis of patterns where the prefixed verb adds an object in the accusative, which were found to be prominent in both Czech and German, adds support to the claims about the analogies between Slavic and Germanic prefixes and verb particles that have been made in the literature. Our approach also makes it possible to quantify and compare the degree to which these effects on valency are productive across both languages, as well as across the different forms of prefixes. Compared to an analysis based on valency dictionaries, this approach is advantageous not only because it allows us to analyze a larger sample of verbs, but also because no a priori decisions about whether a dependent is relevant for a verb's va-

lency characteristics are made – all dependents that are present in the data are taken into consideration, and the prominent patterns emerge from the quantitative analysis.

References

Nadine Aldinger. 2004. Towards a Dynamic Lexicon: Predicting the Syntactic Argument Structure of Complex Verbs. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, pages 427–430, Lisbon, Portugal. European Language Resources Association.

Nicole Dehé. 2015. Particle verbs in Germanic. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen, and Franz Rainer, editors, *An International Handbook of the Languages of Europe. Word-Formation*, pages 611–626. De Gruyter.

Filip Ginter, Jan Hajič, Juhani Luotolahti, Milan Straka, and Daniel Zeman. 2017. CoNLL 2017 Shared Task - Automatically Annotated Raw Texts and Word Embeddings.

Martin Haspelmath and Thomas Müller-Bardey. 2004. Valency change. In Geert Booij, Christian Lehmann, Joachim Mugdan, Stavros Skopeteas, and Wolfgang Kesselheim, editors, *Morphologie*, pages 1130–1145. De Gruyter.

Martin Haspelmath and Andrea D. Sims. 2010. Morphology and Valence. In *Understanding Morphology*, pages 234–264. Hodder Education, London.

Lukáš Kyjánek, Zdeněk Žabokrtský, Jonáš Vidra, and Magda Ševčíková. 2021. *Universal Derivations v1.1*. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Markéta Lopatková, Václava Kettnerová, Eduard Bejček, Anna Vernerová, and Zdeněk Žabokrtský. 2016. *Valenční slovník českých sloves - VALLEX*. Karolinum, Praha.

Andew McIntyre. 2003. Preverbs, argument linking and verb semantics: Germanic prefixes and particles. In Geert Booij and Jaap Van Marle, editors, *Yearbook of Morphology 2003*, pages 119–144. Springer Netherlands, Dordrecht.

Gillian Ramchand. 2004. Time and the event: The semantics of Russian prefixes. In *Nordlyd 32.2*, pages 323–361. CASTL, Tromsø.

Gillian Ramchand. 2008. *Verb Meaning and the Lexicon: A First-phase Syntax*. Cambridge University Press, Cambridge, UK.

Eugenia Romanova. 2006. *Constructing Perfectivity in Russian*. Doctoral thesis, University of Tromsø.

Helmut Schumacher, Jacqueline Kubczak, Renate Schmidt, and Vera de Ruiter. 2004. *VALBU - Valenzwörterbuch deutscher Verben*. Narr, Tübingen.

Barbara Stiebels. 1996. *Lexikalische Argumente und Adjunkte: Zum semantischen Beitrag von verbalen Präfixen und Partikeln*. De Gruyter.

Milan Straka and Jana Straková. 2017. Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 88–99, Vancouver, Canada. Association for Computational Linguistics.

Peter Svenonius. 2004. Slavic prefixes inside and outside VP. In *Nordlyd 32.2*, pages 205–253. CASTL, Tromsø.

Ida Toivonen. 2020. Verbal Particles, Results, and Directed Motion. In Michael T. Putnam and B. Richard Page, editors, *The Cambridge Handbook of Germanic Linguistics*, pages 516–536. Cambridge University Press.

Dieter Wunderlich. 1997. Argument extension by lexical adjunction. *Journal of Semantics*, 14(2):95–142.

Dieter Wunderlich. 2015. Valency-changing word-formation. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen, and Franz Rainer, editors, *Word-Formation*, pages 1424–1466. De Gruyter.

Joachen Zeller. 2001. How syntax restricts the lexicon: Particle verbs and internal arguments. *Linguistische Berichte*, 188:461–494.

Alexander Ziem, Johanna Flick, and Phillip Sandkühler. 2019. The German Constructicon Project: Framework, methodology, resources. *Lexicographica*, 35(2019):15–40.

Zdeněk Žabokrtský. 2022. *Universal Segmentations 1.0*. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.