Kristel Proost / Arne Zeschel / Frank Michaelis / Jan Oliver Rüdiger

MAP (MUSTERBANK ARGUMENTMARKIERENDER PRÄPOSITIONEN)
A patternbank of argument-marking prepositions in German

Abstract Recent years have seen a growing interest in linguistic phenomena that challenge the received division of labour between lexicon and grammar, and hence often fall through the cracks of traditional dictionaries and grammars. Such phenomena call for novel, pattern-based types of linguistic reference works (see various papers in Herbst 2019). The present paper introduces one such resource: MAP (“Musterbank argumentmarkierender Präpositionen”), a web-based corpus-linguistic patternbank of prepositional argument structure constructions in German. The paper gives an overview of the design and functionality of the MAP-prototype currently developed at the Leibniz-Institute for the German Language in Mannheim. We give a brief account of the data and our analytic workflow, illustrate the descriptions that make up the resource and sketch available options for querying it for specific lexical, semantic and structural properties of the data.

Keywords Argument structure; valency; prepositions; constructicography; construction grammar

1. Argument structure

Learning which participants of a given scene must be expressed in a sentence, and how they must be coded formally, is an essential part of learning any language. Within the framework of Valency Theory, argument structure (or: valency) is understood as a lexical property of particular words (notably verbs). Correspondingly, it has usually been covered in specialised dictionaries which list such properties for individual predicate heads (cf. e.g. VALBU 2004; VDE 2004). However, recent years have seen growing objections to a purely word-centred (“lexicalist”) approach to argument structure, since it fails to capture relevant semantic generalisations across the syntactic patterns of different verbs. This has led to a revised view of argument structure patterns such as [SUBJ V OBJ-1 OBJ-2] as meaningful linguistic signs in their own right (“constructions” in the Construction Grammar sense, cf. Goldberg 1995, 2006). In lexicography and grammaticography, this idea has inspired a range of approaches that seek to complement the traditional formats of valency lexicography with construction-based descriptions (“constructicography”, cf. Lyngfelt et al. 2018). Such descriptions are not only of interest for the scientific community in linguistics, but also hold great promise for practical applications in language pedagogy, where they can serve to spell out ‘hidden’ aspects of grammatical meaning for second language learners.

For means of illustration, in the German valency dictionary VALBU, the verb suchen ‘search, look for’ is said to select for a PP headed by the preposition nach ‘after’ that refers to an object which the searching agent strives to obtain. However, the same is true for many other verbs that are not included in VALBU, but also exhibit a ‘strive to obtain’-reading when combining with a nach-PP. For instance, Duden Online gives the following paraphrases for nach etwas tauchen, nach etwas bohren and nach etwas scharren, respectively: ‘tauchend nach etwas.
suchen, etwas zu erreichen, zu finden suchen’\(^1\) (‘to look for something, try to attain or find something by diving’), ‘mithilfe von Bohrgeräten nach etwas suchen’\(^2\) (‘to look for something with the help of drilling equipment’) and ‘scharrend nach etwas suchen, durch Scharren aus der Erde o. Ä. zu fördern suchen’\(^3\) (‘to look for something by scratching, to try to extract something from the earth etc. by scratching’). When such descriptions are considered side by side, it becomes obvious that the shared meaning is more plausibly ascribed to the shared argument structure pattern \([V + \text{nach} + \text{NP}] \sim \text{attempt to attain NP by V-\text{ing}}\) (presumably inherited from the prototype \text{suchen nach} \text{NP}) than to three special (and each time independent) verb readings of \text{tauchen ‘dive’}, \text{bohren ‘drill’} and \text{scharren ‘scratch’}.

In order to represent such generalisations in a suitable kind of linguistic reference work, novel formats of description are required that are no longer organised around an inventory of words (predicate heads), but around an index of the complex constructions in which these words occur (argument structure patterns).

2. The MAP patternbank

MAP is a new patternbank of German that seeks to provide such descriptions for argument structure patterns with a prepositionally marked argument. We focus on descriptions for constructions with the ten most common prepositions with local (i.e., spatial) source meaning (\text{an, auf, aus, bei, in, nach, über, um, von, vor}). In this section, we describe our workflow in devising these descriptions (2.1) and give a brief overview of the web-based resource in which they are provided (2.2).

2.1 Data and workflow

Our descriptions are based on data from the German reference corpus DeReKo, which contains 53bn tokens of (mostly newspaper) texts from Germany, Austria and Switzerland (cf. Kupietz et al. 2010, 2018). For each preposition, we start out by setting up a preliminary semantic classification of its argument structure patterns based on a literature survey and the data from a large prepositional valency dictionary (see section 2.1.1). Next, a random sample of 1m sentences per preposition is drawn and cleaned in a semi-automatic pre-processing procedure (2.1.2). Third, a sample of 10,000 instances is drawn from the remaining data and annotated for a variety of lexical, semantic and grammatical properties. In the course of this process, the initial semantic classification scheme is usually substantially revised (2.1.3). When data coding is finished, we produce the actual descriptions that are contained in the resource, often resulting in further adjustments to pattern boundaries in order to secure a comparable descriptive grain size.

\(^1\) “tauchen” in Duden online. https://www.duden.de/node/180296/revision/419356 (last access: 04-05-2022).

\(^2\) “bohren” in Duden online. https://www.duden.de/node/132381/revision/542512 (last access: 04-05-2022).

\(^3\) “scharren” in Duden online. https://www.duden.de/node/162037/revision/472136 (last access: 04-05-2022).
2.1.1 Preliminary semantic classification

As a first step, we compare existing descriptions of a given target preposition in specialised preposition dictionaries (Kiss et al. 2016; Schmitz 1964; Schröder 1986) and grammars that deal with such constructions in some detail (Helbig/Buscha 2017; Schulz/Griesbach 1982; Weinrich 2005; Zifonun/Hoffmann/Strecker 1997). In addition, we extract all verbal valency patterns from an electronic version of Müller’s (2013) valency dictionary of German prepositions and devise a preliminary semantic classification of these patterns.\(^4\)

Classifications are formulated as paraphrases involving up to three pattern arguments and the pattern-specific semantic relation that holds between these arguments. The patternbank only covers constructions in which the PP does not have a purely local or temporal adjunct meaning, since such uses of the prepositions in question are already well-documented in many dictionaries and grammars. By contrast, grammaticalised uses as a verbal object marker have long been viewed as essentially meaningless in German linguistics (Duden: Die Grammatik 2016, p. 618; Engelen 1975, p. 111; Pittner 1999, p. 50). Although this view is no longer generally accepted, systematic and sufficiently detailed characterisations of the semantic relations that hold between the arguments of such patterns are as yet wanting (though see Hölein 2019 for a recent research monograph that takes some steps towards closing this gap).

Pattern arguments are labeled with the perspectival roles FIGURE (i.e., trajector, located element) and GROUND (i.e. landmark, reference point). Causative variants of such prepositional FIGURE-GROUND relations include an additional EFFECTOR:

(1) a. \textit{Die Firmen} \textit{bohren nach Öl} \textit{GROUND}  
\textit{‘The companies drill for oil’}

b. \textit{Die Regierung} \textit{lässt die Firmen} \textit{nach Öl bohren} \textit{GROUND}  
\textit{‘The government makes the companies drill for oil’}

Pattern meaning: ‘<FIGURE> attempts to obtain <GROUND>’

In this manner, all pattern meanings are represented in a unified format consisting of two or three arguments (FIGURE and GROUND, sometimes augmented with an additional EFFECTOR) and the pattern-specific semantic RELATION (invoked by the predicate expression, if only indirectly sometimes) that holds between these elements (here: ‘attempt to obtain’). Before we identify these patterns in our own data, some further preparatory steps are taken.

2.1.2 Preprocessing

From the initial sample of 1m sentences per preposition, a number of instances are removed ahead of coding. These include:

- Hits in which the putative preposition is actually a verb particle or product of a tokenisation error

\(^4\) In the case of nach, Müller lists more than 500 patterns with verbal heads. Pattern descriptions consist of an entry (e.g., ‘<jemand taucht nach etwas (Dat.)>’ ‘somebody dives for something (dative)’), a meaning paraphrase (‘jemand begibt sich unter Wasser auf der Suche nach dem Genannten’, ‘somebody goes underwater in search of something’) and one or more associated corpus examples (e.g. ‘Damon Edmunds taucht im Haigebiet vor Südaustralien nach Meeresmuscheln’ ‘Damon Edmunds is diving for seashells in the Southern Australian shark area’; Müller 2013, p. 1957). Prepositions that are more frequent (such as in ‘in’) may have more than three times as many patterns than nach in Müller’s dictionary.
Hit in verbless sentence tokens
- Hits in which the preposition forms part of a set of pre-identified multiword expressions that are irrelevant for our purposes (as devised from idiom dictionaries and other phrasological resources as well as n-gram-analyses, e.g. nach wie vor ‘still’)
- Hits in which the preposition is part of a name (e.g. place name, Frankfurt am Main)
- Hits in which the preposition heads a nominal attribute (e.g. Sehnsucht nach ‘longing for’)
- Hits in which the preposition heads the complement of an adjective (e.g. süchtig nach ‘hooked on’)
- Hits that are putatively purely local in meaning, since the preposition combines with a motion or posture verb (e.g. nach Frankfurt fahren ‘drive to Frankfurt’, nach links neigen ‘lean to the left’, cf. section 2.1.1)
- Hits in which the preposition heads a temporal (or other clear) adjunct (e.g. nach 10 Uhr losfahren ‘leave after 10 o’clock, cf. section 2.1.1)

With quite some variability between the different prepositions, these cases usually add up to at least 25% of the original data. These false positives are then removed in order to relieve subsequent coding. From the remaining data, a random sample of 10,000 instances is drawn on which the descriptions in the patternbank are based.

2.1.3 Coding

These 10,000 samples are then coded manually by at least two annotators for the following properties:
- Pattern meaning
- Predicate properties: lexical head, predicate type (verbal, complex, verbal + reflexive, complex + reflexive), predicate diathesis (active, passive, converse), realised verb complements (other than the three pattern arguments), occurrence in a construction with special argument realisation properties (e.g., infinitival constructions, imperatives etc.)
- FIGURE properties: grammatical function of the FIGURE argument (e.g. “subject”), syntactic category of the FIGURE argument (e.g. “subject clause”)
- GROUND properties: prepositional head (e.g. “nach”), governed case of the complement (e.g. “dative”), lexical head of the complement (e.g. “Öl” in an expression like nach Öl bohren ‘to drill for oil’)
- EFFECTOR properties (where relevant): grammatical function of the EFFECTOR argument (e.g. “subject”), syntactic category of the EFFECTOR argument (e.g. “subject clause”)

During coding, preliminary semantic classifications are adjusted as appropriate in order to accommodate new, hitherto non-anticipated usages. Also, it is common for expressions to be grouped and regrouped several times over as new semantic commonalities come to the attention of the annotators in the course of the classification task.

2.2 The resource

The contents of the patternbank are both textual (comprising different types of articles for descriptions on different levels of abstraction, see below) and diagrammatic (comprising

At a later stage, we plan to include these attributive usages in the patternbank, too.
visualisations of different kinds of formal and semantic usage preferences, cf. section 2.2.1). A complex faceted search allows pattern instances to be filtered for a wide variety of semantic, lexical and structural features (cf. 2.2.2).

2.2.1 Contents

Prepositional argument structure patterns can be described on different levels of abstraction. The meaning associated with the verb-preposition combination nach etwas bohren ‘to strive to obtain something by drilling’ in (1), for example, is also evoked by combinations of the same preposition with other verbs: nach etwas kratzen/tauchen equally evoke the meaning ‘strive to obtain’, where the verbs specify the manner in which the FIGURE strives to obtain the GROUND (by scratching or diving, respectively). Slightly similar meanings arise from combinations of nach with schlagen ‘to hit’, werfen ‘to throw’ and treten ‘to kick’ and from syntagms consisting of nach and sich sehnen ‘to yearn’, lechzen ‘to crave’ and dürsten ‘to thirst’. While the former express the meaning ‘strive to reach someone/something’ (cf. Levin 1993; Perek 2015), the latter evoke the meaning ‘strive to experience’. The three meanings could in principle be taken to call for three different patterns, or for only one abstract pattern whose meaning could be paraphrased as ‘to strive for something’.

The patternbank provides descriptions on different levels of semantic abstraction in order to deal with grain size issues of this kind. Low-level descriptions (“pattern articles”) constitute the core of the patternbank and provide an account of individual prepositional argument structure patterns with their specific meaning, attested formal realisations in the German reference corpus DeReKo and typical usage patterns in the data. Mid-level descriptions (“family articles”) capture semantically coherent families of such patterns that can be differentiated from other families with the same preposition but different meanings. Top-level descriptions (“marker articles”) systematise the full spectrum of families with their associated member patterns, list individual patterns not belonging to any overarching family and address the relation of the abstract pattern-specific meanings of the preposition to its local source meaning. These are the most abstract/inclusive descriptions contained in the resource.

As indicated, the patternbank also contains visualisations of quantitative usage preferences of the relevant prepositional argument marker in the investigated sample (10,000 attestations). These include the relative frequencies of different argument structure patterns and pattern families that share the same prepositional argument marker, the relative frequencies of specific syntactic realisations of a given pattern, the relative frequencies of different case realisations of the GROUND argument (where variable) and the relative frequencies of different lexical fillers of the elements RELATION and GROUND.

2.2.2 Query options

A powerful faceted search permits users to identify samples that match a broad variety of semantic, lexical and structural properties. Semantic properties can be filtered with a tag system for categories such as ‘causal’, ‘scalar’, ‘adversative’ etc. Lexical properties include the lemmatised head of the sample’s predicate expression, its predicate type (reflexive like sich erkundigen ‘to inquire’ vs. non-reflexive such as ringen ‘to wrestle’, simple like fragen ‘to ask’ vs. complex such as Frage stellen ‘to ask a question’) as well as the internal constituents of complex predicate expressions (e.g. V+NP.dat for verbnominal predicates such as Frage stellen ‘ask a question’, Feuer fangen ‘catch fire’ etc.). Structural properties include
various aspects of the formal make-up and grammatical marking of the four pattern elements RELATION, EFFECTOR, FIGURE and GROUND. These include the syntactic function of the EFFECTOR and FIGURE arguments, the governed case of the GROUND argument and the diathesis type of the RELATION element (= predicate expression, i.e. ‘active’, ‘passive’ or ‘converse’). Apart from the internal structure of predicate expressions and the grammatical function of their arguments, it is possible to filter for a range of constructions that may alter the canonical argument realisation of a given pattern (such as e.g. various non-finite constructions and subjectless imperatives).

The query options also allow for semantic, lexical and structural properties to be combined such that users may e.g. search for instances where predicates of a given type are used in a specific syntactic construction and/or with a particular diathesis type and/or a GROUND element in the accusative or dative case. Likewise, users may search for instances of patterns with particular semantic properties selecting a tag like ‘causal’ or ‘scalar’ and then filter instances of these patterns for realisations of the EFFECTOR or FIGURE arguments in a particular syntactic function.

In the following section, we give an overview of the structure and contents of pattern-level descriptions (“pattern articles”), which form the backbone of the resource.

3. Pattern articles

Pattern articles contain information on the meaning and form of prepositional argument structure patterns as well as their instantiating predicates. These three types of specifications are described in separate sections of the articles.

3.1 Meaning section

The meaning section gives a paraphrase of the pattern-defining relation between FIGURE and GROUND (i.e., the overall scenario designated by the pattern, e.g. ‘strive to obtain’) and the semantic type of the arguments. If the canonical realisation of the pattern involves an EFFECTOR in addition to the two core arguments FIGURE and GROUND, this is also stated and exemplified in the meaning section.

In case the pattern shows semantic variation, the meaning paraphrase generalises over all meaning variants. Meaning variants are listed in a separate subsection. Variants typically arise from differences in the semantic type of the pattern arguments (entity vs. event or state of affairs, concrete entity vs. abstract entity etc.).

Pattern instances that are ambiguous between two or more meaning variants (Er fragte nach seinem Gehalt ‘He asked for his salary’ vs. ‘He asked about his salary’) are listed and commented on in a separate paragraph of the meaning section. The relation of the pattern-specific abstract meaning of the preposition to its local source meaning is also addressed in a separate subsection (‘conceptualisation’).
3.2 Form section

All formal realisations of a given pattern in the underlying sample (10,000 sentences) are listed in the form section of the pattern articles. Formal realisations are first grouped according to the case of the GROUND element (insofar as there is variability here) and subsequently for their diathesis type (e.g. first all variants with governed accusative case, and within these, first all active variants, then all converse ones, then all passive ones etc.). Within these subsets of expressions, formal realisations are distinguished according to the grammatical function of the FIGURE and EFFECTOR arguments and the structural realisation of the predicate expression. Predicate expressions may be realised as simple lexical units (i.e. lexical predicate heads with all dependencies open, cf. *suchen* ‘to search’) or as (partially) complemented syntags (i.e. lexical predicate heads that have been compositionally combined with one or more of their complements, e.g. *jemanden durchsuchen* ‘to frisk someone’). All lexical complements that are not at the same time arguments of the schematic pattern (i.e. its FIGURE, GROUND or EFFECTOR) are analysed as elements of an “augmented predicate” (i.e. a predicate head in compositional combination with one or more dependents). In contrast to simple predicates (“PRD”), such “augmented predicates” are labelled “PRD+” in the argument structure representation.

For means of illustration, figures (1)–(3) show the representation of three semantically similar, but formally distinct realisations of argument structure patterns with the prepositions *nach*, *um* and *auf*: figure 1 shows an active pattern with two arguments, the marker *nach* and a dative GROUND argument. Figure 2 also shows an active pattern with two arguments, but here with the marker *um* and an accusative GROUND. Figure 3 shows a passive pattern with an augmented predicate and only one realised pattern argument, the accusative GROUND introduced by *auf*:

---

This excludes types that differ from standard realisations only because of construction-independent grammatical regularities (e.g. subjectless variants that owe the missing argument to an infinitival or imperative realisation of the predicate). Instances involving such special constructions (“Sonderformen”) can be identified in the pattern search, however.
3.3 Predicates section

Predicates instantiating a given pattern are listed according to their predicate type (simple non-reflexive verbal predicates, simple reflexive verbal predicates, complex non-reflexive predicates and complex reflexive predicates). The four lists of predicates are accompanied by an annotated corpus example for each element on these lists. In addition to these mere lists, instantiating predicates are grouped into semantic classes, and their interaction with the semantics of the pattern is discussed.

The predicate section closes with a commentary on “shared arguments” that addresses the question of whether the GROUND argument of the pattern is also an argument of particular instantiating predicates. Where this is the case and the lexical argument is also specified as a PP of the relevant type (e.g. nach + dative, as in sich nach etwas erkundigen ‘to ask about something’ in the ERLANGEN-pattern), the GROUND argument is considered as shared by the pattern and the instantiating predicate; it is an argument of both.

In other cases, the GROUND argument may also be a semantic argument of the instantiating predicate but is conventionally realised in a different form, for example as an accusative NP (2.a), as a finite complement clause (2.b) or as a PP with a different preposition (2.c):

(2)  a. Er sucht nach einer Wohnung. / Er sucht eine Wohnung.
    ‘He searches for an apartment.’
  b. Es wird nach dem Sinn von Agentensystemen gefragt. / Es wird gefragt, ob Agentensysteme
      sinnvoll sind.
    ‘People wonder about the purpose of agent systems/whether agent systems make sense.’
  c. Die Kinder betteln nach / um Schokolade.
    ‘The children beg for chocolate.’
In case the alternative realisation is in fact the canonical one (as in betteln um as opposed to betteln nach), attestations in the target construction indicate that this construction is productive, since it encroaches on the semantic territory of a competitor. Where the GROUND argument of the pattern is not an argument of the instantiating predicate, valency approaches analyse it as an adjunct (e. g. nach etwas buddeln ‘strive to obtain something by digging’, nach etwas kratzen ‘strive to obtain something by scratching’ and nach etwas brüllen ‘strive to obtain something by roaring’). From a pattern-based perspective, however, these arguments are said to be ‘coerced’ on the predicate by the schematic argument structure pattern, thus testifying to its independent meaningfulness (Goldberg 1995).

4. Outlook: potential applications

Although the resource is primarily intended as a service to the scientific community in linguistics, previous presentations of the project have indicated that it may be worthwhile to showcase specific possible applications of the patternbank also for other user types once the prototype is publicly available.

One such context is academic teaching. Instructors have pointed out that it would be very interesting for them to provide students with resources that permit a direct, side by side comparison of lexical valency- and pattern-based descriptions of one and the same argument structure construction. For predicates that occur in a particular pattern that are also covered in the German electronic valency dictionary E-VALBU, we therefore provide direct links to the relevant verb (reading) entry in E-VALBU.

A second possible application context is second language learning. Governed prepositions pose notorious problems to learners since whatever semantic content they may have retained is not always immediately apparent (cf. section 2.1.1). A description that helps to uncover traces of their original local meaning and puts a given pattern in semantic perspective (by grouping individual expressions with similar uses) may thus be a welcome addition to the spectrum of existing German language teaching materials. Here, it might be worthwhile to consider a gamification component and devise e. g. some kind of quiz that can serve as an entry point into the patternbank.

Last but not least, the predicate lists in individual pattern articles (alongside the associated corpus samples) also perform a quasi-thesaurus function for expressions that instantiate a given grammatical pattern meaning in the German reference corpus DeReKo. Conceivably, this kind of information could also be of interest for other types of linguistic professionals such as translators.

References


Contact information

Kristel Proost
Leibniz-Institut für Deutsche Sprache (IDS)
proost@ids-mannheim.de

Arne Zeschel
Leibniz-Institut für Deutsche Sprache (IDS)
zeschel@ids-mannheim.de

Frank Michaelis
Leibniz-Institut für Deutsche Sprache (IDS)
michaelis@ids-mannheim.de

Jan Oliver Rüdiger
Leibniz-Institut für Deutsche Sprache (IDS)
ruediger@ids-mannheim.de

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.