Draw *mir* a sheep: A Supersense-based Analysis of German Case and Adposition Semantics

Jakob Prange · **Nathan Schneider**

Abstract Adpositions and case markers are ubiquitous in natural language and express a wide range of meaning relations that can be of crucial relevance for many NLP and AI tasks. However, capturing their semantics in a comprehensive yet concise, as well as cross-linguistically applicable way has remained a challenge over the years. To address this, we adapt the largely language-agnostic SNACS framework to German, defining language-specific criteria for identifying adpositional expressions and piloting a supersense-annotated German corpus. We compare our approach with prior work on both German and multilingual adposition semantics, and discuss our empirical findings in the context of potential applications.

Keywords Natural language understanding · Prepositions · Lexical semantics · Case · German

1 Introduction

German, like many other languages, uses adpositions (such words as *in*, *zu (to)*, *wegen (because of)*, ...) and case (genitive, dative, accusative) to mark specific relations between natural language phrases. And contrary to popular claims (or, none the better, tacit assumptions) in natural language processing

J. Prange
Dept. of Computer Science, Georgetown University,
Washington, DC, USA
E-mail: jp1724@georgetown.edu

N. Schneider
Depts. of Computer Science and Linguistics, Georgetown University,
Washington, DC, USA
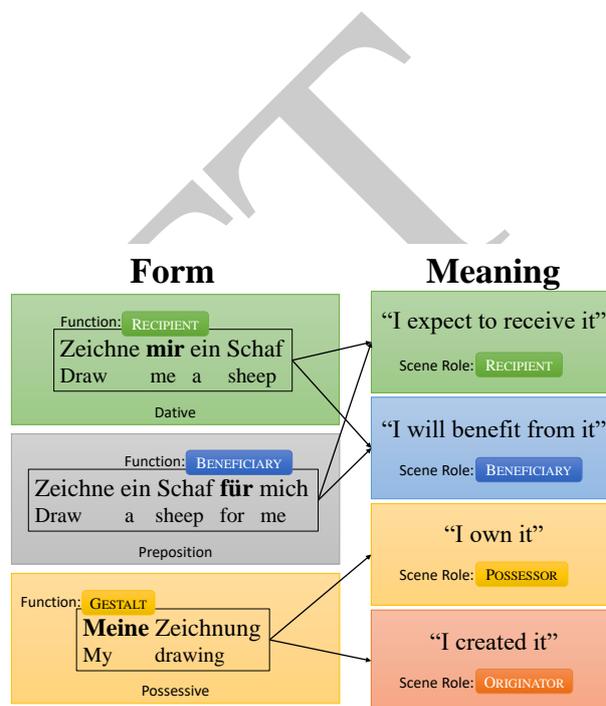E-mail: nathan.schneider@georgetown.edu

Fig. 1: Illustration of adposition and case semantics in German by means of a form-meaning mapping and scene role and function supersenses assigned in our framework.

(NLP) that these relations can or even should be ignored in computational language understanding because they are merely grammatical, they do indeed convey crucial bits of meaning that are worth investigating [61, 51, 71, 65, 17].

To make things more complicated, lexical adpositions (i.e., pre-, post-, and circumpositions) can interact with morphological case markers, as well as each other. For example, a preposition may—under certain contextual conditions—be paraphrased with a different preposition (e.g., *over ↔ above*), a case-marked pronoun (*von mir / of me ↔ mein / my*), or a construction that does not involve any explicit relational signal, such as a noun-noun compound (*Schale für Obst / bowl for fruit ↔ Obstschale / fruit bowl*). But what sorts of patterns arise? What is the influence on the meaning of an utterance or the communicative intent when we choose one adposition, case, or idiomatic construction over another?

Which patterns pertain only to German and which can we observe across languages?

To get a better idea of the meanings and interactions arising from adposition and case usage in German, consider figure 1. The first two sentences are roughly equivalent, despite using superficially diverging constructions (dative case vs. the preposition *für* + accusative case). They can also both be interpreted with respect to two different meaning nuances, one of which involves spatial transfer while the other one does not. The third example involves a possessive pronoun, which can be read as expressing either "true", alienable possession or, alternatively, creatorship.

Since each one of the different adposition and case senses may have different real-world implications, being able to distinguish between them is imperative as we move towards more capable and immersive artificial intelligence (AI).

Many researchers have dealt with semantics of adpositions from various angles and proposed theoretical frameworks [51], dictionaries [52], and corpora [18, 51] to facilitate their computational analysis and processing. Our goals for this article are to highlight the importance of this subfield of natural language understanding (NLU) for a wide range of AI applications, and to contribute guidelines, data, and empirical insights regarding a comprehensive semantic analysis of the German adposition and case systems.

We approach these goals by defining in a concise and replicable manner the scope of linguistic phenomena in German that fall within our notion of *adpositional expressions*, and by adapting and applying an existing annotation framework (the Semantic Network for Adposition and Case Supersenses, SNACS [61]) to a German corpus. Our specific contributions are as follows:

– We introduce a classification taxonomy for the semantics of German adpositions and case markers, building upon the SNACS framework and construal analysis. To the best of our knowledge, this is the most comprehensive and and most practical analysis to date of German adpositions and case "in the wild".
– We present a pilot annotation study to validate the task setup, measure its difficulty, and analyze the distribution of adpositions, case, and their semantics. We also discuss qualitatively a number of special and difficult constructions involving adpositional expressions and compare with SNACS analyses of other languages. The annotated data is available to the community at https://github.com/nert-nlp/German-SNACS.
– We situate our research in the fields of language technology and artificial intelligence, by walking through concrete application scenarios that are prone to errors stemming from improper handling of adposition semantics, and showing how our work can help resolve them.

2 Related Work

The problem of formalizing, marking up, and leveraging the meaning nuances signalled by grammatical markers is not a new one, neither in general nor for German in particular. In this section, we review various approaches from the last five decades, both establishing important background knowledge for the reader to be able to follow our assumptions and revealing the gap in the existing research landscape that we intend to fill. We begin by giving a short history of research on German adpositions and case semantics, before going over the fundamentals of Case Theory and Conceptual Metaphor, which inspire core aspects of our analysis. At that point it should become clear how the abstract and general theories of meaning relate to something as seemingly banal as "stop words". We will then introduce the SNACS annotation scheme, which our work directly extends.

2.1 German Preposition and Case Semantics

When investigating the actual semantics of a specific linguistic class, the most obviously necessary first step is to define that class and to identify the phenomena that fall within it. In the case of adpositions, this task is, however, far from trivial, due to many similarities and overlaps with other parts of speech such as conjunctions (e.g., *und / and*, *weil / because*) and adverbs (e.g., *zusammen / together*, *hier / here*), or more broadly, particles. There is a rich literature on this problem in the Germanic and Romance languages dating back to the early and mid-twentieth century, which has been summarized quite brilliantly by Crössmann [13]. Crössmann concludes that logical and semantic criteria are—for the most part— superior to morpho-syntactic ones for drawing these boundaries, and generally agrees with Jespersen [31], Brøndal [7], and Ljunggren [43] that, based on these criteria, the traditionally separate classes of prepositions and subordinating conjunctions should be clustered together as 'subordinators'.

The semantics of German prepositions have since been examined lexicographically [52], by means of cross-linguistic comparison [17, 57, 14, 48], and in terms of their interaction with case, in particular the *two-way* prepositions, which express different meaning nuances depending on the case of their object [65, 80, 72, 77, 15]. They have also featured in computational analyses, e.g., using vector space semantics [35], and applications such as machine translation [75, 74].

An approach that is closely related to ours is that of Müller et al. [51], who developed a hierarchical annotation scheme containing 37 categories for the disambiguation of German prepositions in context. Multiple categories can be assigned to accommodate usages that systematically activate multiple meaning nuances at the same time [33]. They annotated a corpus of news data, focusing on 22 prepositions that

can engage in determiner-less preposition-noun constructions (PNCs).

2.2 Case Theory and Thematic Roles

Fillmore [19] has argued for a proper notion of functional—in addition to structural—relationships as first class citizens of any grammatical theory. This is because functional grammatical notions such as *subject* and *object* (in contrast to structural ones such as *noun phrase*) exhibit rather systematic correspondences with semantic relationships like AGENT and PATIENT or UNDERGOER called thematic *roles*. Case theory was succeeded by the more fine-grained[1] theory of *frame semantics*, which has been implemented in the FrameNet database [2] as well as the German SALSA corpus [18, 8]. A good middle ground in terms of practicality, domain-independence, and granularity is VerbNet [41, 32], whose thematic role inventory directly inspired the one we use in our semantic analysis. In contrast to case grammar, we take the position (spiritually consistent with FrameNet) that these roles are not atomic meaning primitives but rather high-level, protoype-based generalizations over a wide range of roles in more specific conceptual scenarios.

2.3 Conceptual Metaphor and Construal

Another linguistic theory that drives our investigation is that of *construal* in cognitive linguistics [40, 30, 38], which maintains that linguistic form is not arbitrary, but rather the words and constructions we choose to express any given state of affairs *construe* that state of affairs in a certain way. In order to articulate abstract relations in a target domain, we tend to choose words from a source domain that is grounded in *embodied* (i.e., frequently physically perceived) conceptual primitives. The process of construal can be obvious and intentional, like in carefully chosen metaphors (*they got into a **heated argument** about snakes*, EMOTION is TEMPERATURE), or more subtle and intuitive as in abstract or fictive motion (*she saw him **from** afar*, PERCEPTION is MOTION). With respect to adpositions specifically, linguists often make use of semantic networks encoding the breadth of available meaning nuances and the order in which they can be derived from one another [71, 24, 44]. Trott et al. [70] highlight the relevance of construal for the field of NLP, and establish a set of meaning dimensions along which the general phenomenon can be classified for practical applications.

---

[1] For example, the verb *zeichnen* 'to draw' may evoke the specific frame CREATE_PHYSICAL_ARTWORK with the frame elements CREATOR and REPRESENTATION, which are typically instantiated by the syntactic subject and object, respectively.
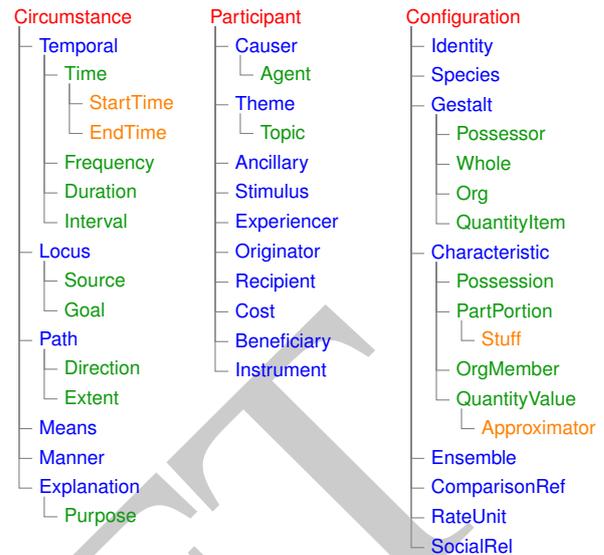

Fig. 2: Hierarchically organized inventory of 50 SNACS supersenses. Replicated from Schneider et al. [62].

2.4 SNACS

The concrete framework we base our analysis on is the Semantic Network of Adposition and Case Supersenses (SNACS) [58, 27, 61], which consists of 50 hierarchically organized categories based on VerbNet thematic roles [32]. The SNACS inventory (figure 2) is broadly divided into senses pertaining to CIRCUMSTANCEs, PARTICIPANTs of scenarios or scenes, and CONFIGURATIONs between entities.

SNACS is designed with typological universals and cross-linguistic applicability in mind. In addition to the English STREUSLE corpus [59, 61], it has so far been adapted to and annotated in Mandarin Chinese [53], Korean [28], and Hindi [1]. Preliminary invesigations are ongoing for Latin, Finnish, French, and Hebrew.

3 Annotation Scheme

We perform our analysis by means of annotating naturally occurring text. Our main assumptions are based on the ones described in the original SNACS guidelines for English [62]. This includes the inventory of categories (figure 2), the construal analysis, according to which annotations are split up into a *scene role* and lexical *function*, as well as the general annotation guidelines for the scene role, which is expected to be stable across translations. The high-level idea is to identify adpositional targets in the text, disambiguate the meaning they express in their respective context (the scene role), and explicate how this meaning is lexically construed (the function). This is illustrated in figure 1.

When it comes down to the details, however, we need to define and handle a range of non-trivial phenomena that pertain to German in particular. We aim for an annotation scheme that is as comprehensive as possible, i.e., we are interested in all instances of prepositions, postpositions, and circumpositions—without cherry-picking, and yet with criteria for including certain word types and morphemes that may not traditionally be considered adpositions or for excluding ones that do not have a sufficiently independent semantics.

Furthermore, German is a fusional language, marking case morphologically in a way that is not clearly and unambiguously separable from lexical items (contrast, e.g., Korean, where case markers are always transparent suffixes, and English and Chinese, which have essentially no grammatical case). Since case may, depending on context, determine its governing adposition's meaning nuance, or carry its own semantics—which, in turn, may typologically correspond to a lexical adposition in another language—we want to analyze it in the same framework, to the extent possible.

Here we describe in detail the policies we have devised.

3.1 The Scene Role/Function Distinction

In our annotation, we apply the *Construal Analysis* as proposed by Hwang et al. [27]. This means that each annotation target is assigned up to 2 different semantic categories, the first one denoting the adpositional object's role within the scene described in the clause (the *scene role*, cf. §2.2) and the second one denoting the prototypical meaning of the adposition or case marker, independently of the current context (the *function*, cf. §2.3). Both scene role and function labels are drawn from the SNACS inventory (figure 2). The construal analysis is illuminating when the scene role conveyed in the current context diverges from the prototypical function(s) of the annotation target. We write, e.g., POSSESSOR↝LOCUS and say "POSSESSOR construed as LOCUS" (e.g., *ich trage einen Schirm **bei mir*** 'I am carrying an umbrella **with** (lit. **at/on**) me'). If the meaning conveyed in the current context is also a prototypical meaning of the annotation target, scene role and function are congruent. For simplicity, we write only a single category in this case (e.g., TIME rather than TIME↝TIME), but annotators are always asked to make the function explicit even if it is congruent with the scene role.

When choosing the scene role, we consider the type of semantic relation the adpositional object engages in with respect to the governing scene, predicate, or entity in each specific context. This abstracts from surface forms and is expected to remain stable across paraphrases and translations.

The function, on the other hand, is tied to the concrete lexical item or construction chosen to express the abstract relation. Both the overall distribution of functions and which function label is chosen for a given target in context are thus sensitive to lexical, syntactic, and cross-linguistic variation.

Deciding what constitutes the set of prototypical functions for an adposition is not trivial. For annotation in German, we settled on the following order of priority:

1. We compiled an open-ended lexicographical list of individual adpositions and their possible functions, along with detailed examples, which should be consulted by annotators before moving on to the more general rules.
2. For prototypically spatial and spatio-temporal adpositions,[2] the function should always be spatial or temporal (i.e., from the LOCUS, PATH, or TEMPORAL sections of the hierarchy). Temporal meanings are generally considered as prototypical as spatial meanings, i.e., we never annotate a spatial function for a temporal scene role or vice versa. If the scene role is spatio-temporal as well, the function should match the scene role. Otherwise, we prefer to choose as the function the spatial category that most adequately describes the metaphor or origin of the extended meaning, except if a temporal category is clearly more salient.
3. Possible functions for genitive case and possessive pronouns are GESTALT and POSSESSOR, following Schneider et al. [62]. If and only if the scene role is POSSESSOR, i.e., the meaning in context is that of true, alienable possession, the function is also POSSESSOR. For all other scene roles, from general association meanings (e.g., GESTALT) to specific extended senses (e.g., ORIGINATOR or SOCIALREL), the function is GESTALT.
4. The function label for dative case and pronouns is always RECIPIENT, regardless of the scene role.
5. Since we aim for a comprehensive analysis of adpositions rather than a limited set of forms, it is possible that an annotator encounters a form that is not yet documented in our framework. In this case, the target will be examined and discussed by the project leaders and eventually added to the guidelines.

3.2 What Counts as *Adpositional*?

As tempting as it may be to simply assume that the terms "adpositions" and "case markers" are inherently well-defined, the question of which linguistic units we ought to include in our semantic analysis is far from trivial. One can draw from a wide range of morphological, syntactic, and semantic criteria in making this decision, and there is no one way that does not make any compromises. It all depends on what the goals of the investigation and their potential applications are.

In line with our main goal—broad-coverage exploration of the mapping between certain relational semantics and linguistic form, both within and across languages—we follow Crössmann [13], Huddleston and Pullum [26], and Schneider et al. [58, 59, 60, 61] in using a mix of syntactic and semantic criteria for the inclusion of linguistic units. More concretely,

---

[2] These are mostly closed-class; a list is provided to annotators.

we consider an expression adpositional and thus markable if and only if it denotes an **asymmetric, or directed figure–ground relation between two concepts**. This is in stark contrast to most prior annotation work, which typically focuses on a limited subset of forms [51, 33, 48]. Further, any form that can be used as a classic preposition (with an NP complement) is always considered an annotation candidate, even when it takes a non-NP complement (e.g., a clause) or no complement at all.

On the other hand, we want to formulate our task—at least for the time being—as strictly *lexical* disambiguation for practical reasons, thereby excluding from being annotation targets any bound morphemes that cannot be clearly separated from their lexical head, as well as any implicit or null-instantiated relations.

As a result, we annotate a range of linguistic units, from "classic prepositions" to prepositional adverbs (or pronominal adverbs), verb particles, multiword adpositional expressions, comparative conjunctions, purpose infinitivals, as well as possessive and dative pronouns. We discuss each phenomenon in detail below.

3.2.1 Classic Prepositions (PREP)

Unsurprisingly, the definition given above includes those words traditionally called prepositions, like *in, auf, über, nach,* and *zu*. They have in common, most notably, that they are (usually) morphological primitives and that they (usually) govern the case (accusative, dative, or genitive) of their object NP, which they precede. Semantically, they establish a relation between their syntactic governor as the more prominent trajector or figure, and their syntactic object as the anchoring landmark or ground.

Centered around this core set of "classic prepositional" word forms, we also include the following special cases, which are somewhat divergent in their morpho-syntactic properties but still share the same semantics:

- **Postpositions**: Some prepositions, such as *nach* 'after, according to' and *wegen* 'because of', can also follow their object rather than preceding it. Apart from word order, these postpositions behave in the same way as prepositions and are thus included. We use the term *adpositions* to emphasize this.
- **P's with non-NP complements**: We annotate prepositional word forms, even when they are used with a non-NP complement. This could be, e.g., adjectives (*etw. **für** gut halten* 'deem sth. (**for**) good'), or VPs/clauses, in which case the marker would traditionally be assigned the part-of-speech of 'subordinating conjunction' (SCONJ; ***bis/als*** *wir ankamen* '**until/when** we arrived').
- **P's with implicit complements**: While, in English, canonical prepositions can also occur without any explicit object ('intransitive prepositions' [61, 26]), this

kind of construction in German typically requires the preposition to undergo some form of morphological change into either pronominal adverbs or verb particles, which are discussed below in §3.2.2 and §3.2.3, respectively.

- **P+D contractions**: When immediately followed by a determiner (usually a definite article), certain prepositions may (or even *must* in fluent, natural German) contract with that determiner to create a special form (*zu + der → zur, an + dem → am*). While there are constraints on when this can happen, and the contracted forms may have special semantics and/or pragmatics [56, 25, 76, 11], these properties are finer-grained than what we aim to capture. We split them up in preprocessing and annotate the preposition part in the same way as uncontracted ones.

3.2.2 Pronominal Adverbs (PAV)

Out of the multiple ways a German preposition can undergo a morphological process to end up intransitive (i.e., argument-less) is by combining with a locative particle (*da, wo*), as in the following examples:

- *da* 'there' + *für* 'for' → *da**für*** 'for that'
- *wo* 'where' + *mit* 'with' → *wo**mit*** 'with what/which'

The resulting word class is called *pronominal adverbs* or *prepositional adverbs*. These are deictic expressions, encoding both the prepositional marker and a reference to the prepositional object in a single word form. The object referred to is usually an established discourse entity. We include in this class combinations of prepositions + emphatic demonstrative articles (*des**wegen*** '**because of** this', ***trotz**dem* '**in spite of** this'), and prepositions + reciprocals (***mit**einander* '**with** one another'). Additionally, we annotate all word tokens that are homographs with a pronominal adverb, even when they are used as clausal subordinators (e.g., *damit* with its purposive sense of 'so that').

3.2.3 Verb Particles (VPRT)

Another interesting class of words involving adpositional expressions is that of the particle verbs (PVs). Here, a verb combines with an adposition (or adposition-derived formative) to form a new verb whose semantics, and sometimes argument structure, is somewhat different from the base verb. This change in meaning can be more transparent, i.e., $[\![PV]\!] \approx [\![V]\!] + [\![P]\!]$, or more opaque.

This is, however, not a clear-cut distinction that can always be easily made, but rather lies on a continuous scale. An example for a highly transparent PV is ***auf**stehen* 'to stand/get **up**', a slightly less transparent one is ***auf**wachen* 'to wake **up**', and a highly opaque one is ***unter**halten* '**under**+hold = to entertain'.

Orthogonally, German PVs can be divided into *trennbar* (separable) and *nicht trennbar* (non-separable). In V2-ordered main clauses, separable PVs leave their adpositional part in the clause-final position when moving[3] the inflected verb part to V2, whereas non-separable ones move as a whole.

Because the dimensions of semantic transparency and morpho-syntactic separability do not align, we apply a rather crude preprocessing method to ensure high recall of our semantic annotation: we manually separate all PVs whose adpositional part matches a regular preposition (this does not include PVs beginning with, e.g., *hin-* or *her-*), make it available for annotation, and rely on annotators to determine its semantic content, if any.

3.2.4 Multiword Adpositional Expressions (MWE)

While we do not currently attempt to analyze subword units (except when they can be unambiguously isolated, as discussed above), annotation targets need not consist of only a single token. German has several constructions that we consider to be multiword adpositional expressions:

– **Purposive *um zu* infinitivals** (INF): This combination of the preposition *um* and the infinitival marker *zu* expresses the purpose of an action, similar to the English *(in order) to*. Potential direct and indirect objects as well as adjuncts of the complement VP intervene between *um* and *zu*, resulting in a discontiguous annotation target (***um_ es _zu*** finden '**(in order) to** find it'). Similarly, when its complement is a separable PV, *zu* is embedded between the adpositional and verbal parts (***um_ auf_zu***stehen '**(in order) to** get up'). The two can occur together, leading to even larger gaps (***um_*** es auf_**zu**heben '**(in order) to** pick it up).

– **Circumpositions**: These two-part adpositions surround their object and are thus always discontiguous (***von_ Kindesbeinen _an*** 'ever since childhood', ***um_ des Geldes _willen*** 'for money's sake'). The second, postpositional part is usually derived from an adposition (*an*) or a nominal (*willen*) itself. Most circumpositions are not in frequent and productive use anymore and, in fact, none of them are attested in our data.

– **Genitive P's with *von***: Certain prepositions governing genitive case can instead combine with *von* 'of' (which then governs dative case), with little to no change in meaning (***innerhalb_von*** 'within') [63, 66, 10, 16]. We consider this a superficial variant of the main preposition rather than two separate annotation targets.

– **Idiomatic P+P combinations and PNCs**: Following Schneider et al. [61], we also include a number of lexicalized, highly idiomatic expressions consisting either of multiple prepositions (*nach_und_nach* 'little by little',

bis_auf in the sense of 'except') or of prepositions and incomplete nominals or adverbs (*auf_gut_Glück* 'haphazardly', *von_Zeit_zu_Zeit* 'from time to time', *auf_einmal* 'suddenly').

Yet other types of frequently collocated prepositional expressions or phrases are semantically transparent enough that we annotate each adpositional constituent separately, as described in §3.2.1, rather than the expression as a whole:

– **Transparently stacked P's**: In contrast to the idiomatic *bis_auf* meaning 'except' as pointed out above, *bis* 'until' can also compositionally stack with a number of spatial (GOAL-type) preposition like *zu* 'to' and *an* 'at/on(to)'. Here, the respective second GOAL preposition retains its usual meaning, while *bis* profiles the EXTENT of the PATH taken to reach the GOAL. We therefore annotate both prepositions separately.

– **Conventionalized transitive PPs**: German has many other conventionalized PP types which, because their nominal part often has relational semantics and thus takes a genitive or PP complement itself, in a way act as complex prepositional units (*im Laufe* 'in (the) course (of)', *mit Hilfe* 'with (the) help (of)', *in Folge* 'following/as (a) result (of)'). There are certainly arguments to be made that these expressions are on the verge of being grammaticalized, perhaps most tangibly by means of the contracted orthographical variants *mithilfe* and *infolge*. However, since it is still possible to derive their meaning compositionally, we treat them as normal, transparent PPs in our annotation.

3.2.5 Comparatives: als, wie, and zu

Since comparative constructions fall within the range of our COMPARISONREF and EXTENT senses, we include the comparative markers *als* 'than', *wie* 'as/like', and *zu* 'too' in our annotation, even though they are not traditionally recognized as prepositions. *Als* and *wie* are often called 'comparative conjunctions' and can otherwise also function as the adverbial subordinators 'when' and 'how', respectively. We follow our policies established above in annotating all of these adpositional, i.e. asymmetric-relational, usages that share a word form with a classic NP-complement preposition.

3.2.6 Possessive (POSS) and Dative (DAT) Pronouns

Motivated by the observation that, cross-linguistically, morphological case shares much of the semantics of adpositions—in fact, adpositions can be seen as a finer-grained and more flexible extension of the case system [19]—we want to include case markers in our annotation, whenever they are easily accessible as lexical items.

The practice to annotate possessives has a precedent in English SNACS annotation, as piloted by Blodgett and

---

[3] In a purely illustrative sense. We are agnostic with respect to the formal syntactic notion of movement or any other syntactic paradigms.

Schneider [5], who found, e.g., that the meaning relations signaled by possessive constructions are often much broader and more abstract than just concrete alienable possession. The English possessive markers are *of*, the clitic *'s*, and possessive pronouns, all of which are 'lexical' in the sense that they can easily be tokenized and annotated. Possessive, and generally case-marked pronouns are similar to pronominal adverbs (§3.2.2) in that they encode both the case marker and a reference to its object within a single form. Due to German fusional morphology, genitive case marking on articles and nominals is not as easily accessible as *'s* in English. Thus, apart from the preposition *von* 'of', we only include possessive *pronouns*.

By the same logic, we are interested in German dative case, whose semantics we anticipate to span a variety of concrete and abstract TRANSFER, PERCEPTION, POSSESSION, as well as general ASSOCIATION frames [44]. To our knowledge, this is the first empirical investigation of its kind to cover the semantics of the German dative together with possessives and adpositions, all within the same framework. We analyze the arising form-meaning interactions in §4.6.1.

3.3 What does NOT Count as *Adpositional*?

Conversely, we decided a priori on several phenomena we do not aim to capture in the current version of our framework. For the most part, these decisions are again based on semantic criteria, i.e., we rule out specific kinds of case marking with systematically less salient or even rather bleached semantics. This includes any nominative and accusative case marking, certain reflexive pronouns, any case governed directly by a lexical adposition, and any adjectives and participles (even when their semantics resembles that of an adposition).

3.3.1 Nominative and Accusative Case

Subjects and direct objects, marked in German with nominative and accusative case, are not annotated in the current version of our framework for the main practical reason that they are much more frequent than adpositions and would substantially increase annotation effort. However, as Shalev et al. [64] demonstrated for English, the SNACS senses can in principle be applied to core-syntactic arguments. We plan to adopt this for German in future work.

3.3.2 Reflexives

The syntactic argument structure of certain verbs demands reflexive objects, i.e., pronouns in accusative or dative case that corefer with the subject (e.g., *sich* 'oneself'). While we do not currently annotate any accusatives, as explained in the previous section, we distinguish two types of dative reflexives: those that denote an additional semantic relation between the subject and the verb, and those that are purely grammatical. We include the former but not the latter. To test which class a given dative pronoun belongs to, we ask annotators to try to substitute the pronoun with a non-reflexive one (e.g., *er kocht **sich** etw.* vs. *er kocht **ihr** etw.* 'he is cooking sth. **for himself/her**'). If the result is grammatical and the meaning still revolves around the same type of event, just with a different participant, the target is included. Otherwise (e.g., *er stellt es **sich** vor* 'he imagines it' vs. *er stellt es **ihm** vor* 'he introduces it **to him**'), it is discarded.

3.3.3 Adposition-governed Case

Most German prepositions govern the case of their object. Some always take objects of the same case (accusative, dative, or genitive), and others are more flexible. Dubbed 'two-way prepositions' and generally situated in the locative semantic domain, these markers can either combine with datives to denote a stative location, or with accusatives to denote the endpoint of a (not necessarily physical [65]) motion or change. The case governed by a particular two-way preposition in a specific context can thus be used by annotators to disambiguate its semantic function (either LOCUS or GOAL) and evoked scene role. However, we do not consider adposition-governed morphological case markers annotation targets in their own right.

3.3.4 Adjectives and Participles

As an exception to our strict adherence to semantic inclusion/ exclusion criteria, we also rule out adjectives and participles although they sometimes signal relations that overlap with adpositional semantics and may even govern case. The criterion for ruling out such an item is that it still follows, in a productive way, the morpho-syntactic behavior of its respective part-of-speech class (adjectives can be compared, verbs that participles are derived from can be conjugated).

4 Annotated Corpus

In order to test out the framework and guidelines as well as empirically evaluate their quality and difficulty, we annotated a pilot corpus of natural language data. Annotation was carried out over multiple rounds, with two different sets of annotators, and in reciprocal alternation with refinements to the guidelines based on empirical analysis of the annotation quality and feedback from the annotators.

Throughout all iterations, annotators were presented with continuous stretches of text containing adpositional and case markers naturally embedded in context. The annotation targets had been manually identified by the first author of this

paper, according to the criteria outlined in §3.2 and §3.3, and highlighted in the text for annotators to disambiguate.[4]

Below we provide details on the data, annotators, and inter-annotator agreement; discuss our overall findings as well as particularly interesting or difficult observations in depth; and point out crucial parallels and divergences between the case-adposition systems in German and other languages, based on a few selected examples.

4.1 Data and Preprocessing

We use a publicly available German translation of Antoine de Saint-Exupéry's *Le Petit Prince* ("Der kleine Prinz", translated from French by Romy Strassenburg, 2015, BUCH-FUNK Verlag). The top-left portion of table 1 provides basic statistics. We choose this text for its availability in many languages and frequent appearance in linguistic and computational semantics research [e.g., 68, 47, 3, 4], and expect the SNACS-annotated German version to contribute to future comparative analyses and multilingual applications alike. It should be noted, though, that parts of the vocabulary and structure of this corpus are noticeably specific to the literary genre and the data can therefore not be considered representative of German as a whole. Moreover, translations (as opposed to text originally authored in a language) are known to exhibit certain tendencies such that 'translationese' is a distinctive style [36]. We thus hope that the methods piloted on *Der kleine Prinz* will be repeated for German texts covering a wide range of genres and styles to better reflect the language as a whole.

Our preprocessing steps were as follows: We first automatically extracted the raw text from the pdf; then tokenized it with Stanford CoreNLP [46], version 3.8.0; normalized abbreviated and article-contracted prepositions using lookup tables; and finally cleaned up chapter headings and page numbers.

4.2 Training of Annotators

First, annotators were sensitized to the class of word forms and meanings introduced in §3 (e.g., by asking questions like "In your own words, how would you describe the meaning of adposition p in sentence X? How is it different from the meaning of p in sentence Y? How is it different from preposition q?"). Then the general annotation procedure with SNACS was demonstrated by means of individual and relatively simple examples. Finally, annotators were asked to

	v0.1	v0.2	LexCats			
Chapters	27	5	(of 1,586 ann. v0.1 targets)			
Sentences	1,509	228				
Tokens	18,600	2,034	PREP	836	PAV	110
Candidates	1,715	244	POSS	215	SCONJ	59
Annotated	1,586	236	DAT	183	INF	31
Unique forms	184	59	VPRT	143	MWE	9

Table 1: General statistics of our corpus. The discrepancy between *Candidate* and *Annotated* targets is explained in §4.4.

self-learn how to apply the written guidelines to sample data. This was complemented with feedback from the first author (= annotator A) and discussions among the annotators.

4.3 Details on Annotators and Annotation Procedure

The annotation process so far has been divided into two main stages with distinct sets of annotators and different versions of the inventory and guidelines, which we call v0.1 and v0.2:

v0.1   Four trained annotators (A, B, C, and D) annotated all 27 chapters of the *Little Prince* in paid work over five months (January – May 2019). All four are fluent in German, though only A and B are native speakers. A and D have advanced training in computational linguistics and B is experienced in teaching German. Each adpositional target was annotated by two annotators independently, and disagreements were resolved (adjudicated) by a third, impartial annotator. There was no full-fledged German-specific guidelines document yet, so the English/general guidelines v2.0 [60] along with lists of German examples were used. Annotators were asked to translate the relevant text snippet into English and look for a matching or closely related example in the guidelines. The translation did not need to be literal, but it was required that the adposition's object and governor also be mentioned in the English translation and reflect the same scenario and role as in the German sentence.

These instructions are obviously quite complicated and demanding, leading to rather low agreement among the annotators (∼40% average raw pairwise agreement on both scene role and function). We therefore made several changes to the workflow and started developing detailed German-specific guidelines. Independently, the SNACS sense inventory and general guidelines were updated multiple times [64, 29]. In order for our analysis to reflect this more current and reliable version of the framework, we conducted another round of annotation.

v0.2   Three trained annotators (A, E, and F) annotated five chapters (1–3, 6–7) in voluntary, unpaid (E and F) / paid (A) work over one month (August 2020). All three are native German speakers and computational linguistics students. Each target was annotated by all three annotators independently in three explicit steps: 1. identifying the semantic argument and head of the adposition, 2. assigning the argument's scene

---

[4] In order to efficiently gather supersense-annotated data on a larger scale, the target identification step may be automated. Schneider et al. [61] have demonstrated for English that this is possible, at least approximately, based on part-of-speech tags and syntactic dependencies. We expect their method to be applicable to German as well, but this still needs to be tested empirically.

	N	Avg. Pairwise IAA		≥ 2 Agree	
		Role (κ)	Fxn (κ)	Role	Fxn
Exact	47	.51 (.50)	.70 (.64)	.87	.96
Depth-2	24	.59 (.61)	.80 (.76)	.90	.99
Depth-1	3	.76 (.64)	.90 (.82)	.99	.99

Table 2: Inter-annotator agreement rates in a v0.2 sample (chapters 6 and 7) consisting of 95 adposition tokens (28 unique lemmas). We show both averages of pairwise agreement (raw and Cohen's κ [12]) and the proportion of targets for which at least (any) two annotators agree, at different levels of supersense granularity (*N* is the number of available sense types at each level).

LexLemma	Approx. transl.	#	unique...		
			Forms	Roles	Fxns
POSS		42	14	7	2
DAT		30	5	6	1
in	'in, into'	19	1	6	3
von	'from, by, of'	16	1	6	1
auf	'on (top of), onto'	14	1	6	3
als	'than, when, as'	13	1	4	3
mit	'with'	12	1	7	4
an	'on, at'	10	4	6	3
zu	'to, too'	9	2	5	3
wie	'like, as, how'	9	1	4	1
Total		**236**	**59**	**31**	**25**

Table 3: 10 most frequent adposition types in the v0.2 portion of our German corpus.

role with respect to the head, 3. assigning the adposition's function. A German-specific guidelines document (in statu nascendi) was used, containing detailed explanations and examples of functions of German adpositions, and referring to the English/general guidelines v2.5 [62] for scene roles, which are expected to transfer across languages. This procedure resulted in much better agreement.

In table 1 we give an overview of annotation targets—broken down by morpho-lexical categories (LexCats, as defined in §3.2)—in the full corpus, most of which has not been updated to v0.2 yet. In the detailed analyses below, we focus on v0.2 as it uses the most current version of the SNACS inventory and annotations are more reliable.

4.4 Governed, Opaque, and Unresolved Targets

It turns out that not all of the targets made available for annotation carry equal amounts of semantic weight. At the lower end of this spectrum are two particular phenomena that are intentionally not fully captured by our current scheme:

- **Governed usages**: These are core arguments of verbs whose adpositional marking does not add any additional semantics. We annotated them as THEME (64 in v0.1, 7 in v0.2).
- **Opaque targets**: In order to achieve high recall, we make all verb particles and dative pronouns available for annotation and rely on annotators to filter out the false positives, such as semantically bleached VPRTs and purely grammatical reflexive DATs (67 in v0.1, 8 in v0.2).

Additionally, 62 targets in v0.1 could not be resolved even after adjudication. Opaque and unresolved cases are excluded from the set of annotated targets in the remainder of this section.

4.5 Inter-annotator Agreement

We show different measures of agreement between annotators in table 2. We observe average Cohen's κ values of .50 on the scene role and .64 on the function ("moderate" according to Landis and Koch [39]), which highlights the difficulty

of the task and suggests that there are still improvements to be made to the scheme and/or annotator training. The hierarchical structure of SNACS allows us to conflate supersenses at different levels of granularity. Agreement is much higher with a coarsened sense inventory (up to .64, .82 with fully flattened hierarchy), showing that annotators generally understand the CIRCUMSTANCE/PARTICIPANT/CONFIGURATION trichotomy and disagreements tend to be on finer-grained and thus more difficult distinctions.

In the vast majority of cases, 2 out of 3 annotators agree. This is a good indicator for the general validity of our scheme. The discrepancy between this and pairwise agreement measures suggests minor but systematic individual differences between annotators' interpretative tendencies, such that any one pair of annotators may have consistent disagreements.

Overall, our agreement numbers are comparable with those of Shalev et al. [64] (English, Wikipedia) and Müller et al. [51] (German, newswire), but substantially lower than the ones reported in Schneider et al. [61] (English, *Little Prince*) and Peng et al. [53] (Chinese, *Little Prince*). Potential reasons for this might be different levels of annotator expertise as well as logistical details and available resources in carrying out the annotation. Additionally, it is not inconceivable that the set of linguistic phenomena we include (e.g., datives, verb particles) or even certain properties of the German language itself (e.g., attachment ambiguity, inflection) are genuinely more difficult to disambiguate.[5]

4.6 Analysis

We show the most frequent adposition types among the v0.2 corpus in table 3. Under 'LexLemmas' we conflate

---

[5] The moderate agreement also suggests that the task in its current form is likely too difficult for crowd workers without either a background in linguistics or close supervision by a project manager. Alternative task formulations that facilitate crowdsourcing have been proposed by Gessler et al. [20].

German				English				Chinese			
Role	%	Function	%	Role	%	Function	%	Role	%	Function	%
LOCUS	11.4	LOCUS	17.5	LOCUS	11.7	GESTALT	15.0	LOCUS	31.6	LOCUS	34.7
MANNER	6.8	RECIPIENT	12.9	PURPOSE	6.6	LOCUS	14.6	TIME	10.6	TIME	9.2
EXPERIENCER	6.8	GESTALT	12.5	TIME	6.5	PURPOSE	7.4	THEME	7.9	DIRECTION	8.4
TOPIC	6.4	SOURCE	7.9	POSSESSOR	5.9	GOAL	7.0	RECIPIENT	7.8	THEME	7.9
TIME	6.4	GOAL	6.7	GOAL	4.3	POSSESSOR	6.5	EXPLANATION	5.9	EXPERIENCER	6.3

Table 4: Most frequent supersense types in our German corpus, the English STREUSLE corpus, v4.3 [61], and the Mandarin Chinese Little Prince corpus [53]. Shaded cells indicate language-specific functions.

all morpho-lexical variants of a single adpositional lexeme ('unique Forms' column, e.g., inflection of POSS and DAT pronouns; PAV and VPRT versions of PREPs). All of the listed target types fill various different semantic roles, depending on context. Interestingly, LexLemmas that are traditionally classified under multiple syntactic categories, like *als* and *wie*, appear to be less semantically ambiguous (4 unique roles across 2 main branches: CIRCUMSTANCE and CONFIGURATION) than, e.g., POSS and *mit* (up to 7 unique roles across 2–3 main branches). Most lemmas can also bear different lexical functions.

In the left part of table 4 we list the most frequent supersenses, separately for scene role and function annotations. Next to LOCUS and TIME, the adpositions in our corpus frequently mark the MANNER (*wie, in, als, auf, ohne*), EXPERIENCER (DAT, *für*), and TOPIC (*über, von*) of a scene.

4.6.1 Datives

Our inclusive approach to analyzing dative semantics together with adpositions and possessives in a single shared framework allows us to gain deeper insights into where the German dative overlaps with other forms and meanings, as well as empirically confirm theoretical findings about case polysemy. This is a novelty, both compared to previous applications of SNACS and in the German-specific literature.

Almost all of the 30 datives annotated in v0.2 of our corpus are split relatively evenly between the RECIPIENT (12), EXPERIENCER (9), and BENEFICIARY (6) scene roles. Examples are given below.

(1) Sie   haben **mir**:RECIPIENT geantwortet
     they have   me.DAT   responded
     'They responded to me'

(2) Dabei half   **mir**:BENEFICIARY↝RECIPIENT die Geographie
     at.that helped me.DAT   the geography
     'Geography helped me with that'

(3) So absurd es **mir**:EXPERIENCER↝RECIPIENT erschien
     as absurd it me.DAT   appeared
     'As absurd as it seemed to me'

This semantic domain overlaps with those of other forms, most prominently *für* (BENEFICIARY, EXPERIENCER), *zu/an* (RECIPIENT), and possessives (EXPERIENCER):

(4) so sagte ich schlecht gelaunt **zu**:RECIPIENT dem Männchen
     so said I   bad   tempered to   the little man
     'so I told the little man, in a bad mood'

(5) Dabei ging   es **für**:EXPERIENCER↝BENEFICIARY mich
     at.that revolved it for   me
     um   Leben und Tod
     around life   and death
     'This was a matter of life and death for me'

(6) **meine**:EXPERIENCER↝GESTALT Panne
     my   hitch
     'the failure [of my plane]'

These empirical findings are in line with Malchukov and Narrog's [44] typological work on case polysemy. They point out that the dative has historically evolved from benefactive and directional cases. In argument structures corresponding to causal chains, such as transfer, the dative or indirect object is usually situated at the literal 'receiving end' of the chain. Haspelmath's [24] semantic map for the dative domain thus includes DIRECTION, PURPOSE, RECIPIENT, EXPERIENCER, BENEFICIARY, the *dativus iudicantis* (e.g., *das ist mir zu warm* 'this is too hot **for me**', subsumed by EXPERIENCER in our scheme), as well as external and predicative POSSESSORs (e.g., in Russian and French).

Many languages (other than German) feature differential object marking, where a single 'dative-accusative' case is used for both RECIPIENTs and PATIENTs. This is the case, e.g., in English pronoun morphology, in alternation with the *to*-dative:

(7) a. she gave **me** the drawing
     b. she gave the drawing **to me**

Similarly, syncretisms between directional and (stative) locative adpositions can be found in French and several Altaic languages, or between dative and genitive case in Australian and Austronesian languages [44].

4.6.2 Cross-linguistic Comparison

In table 4, we also contrast the head of the supersense distribution in our corpus with SNACS analyses of English [61] and Chinese [53]. Unsurprisingly, LOCUS and TIME are consistently in the top-5 scene roles across all three languages.

LOCUS and similar senses such as GOAL and DIRECTION are also frequent function values cross-linguistically, but otherwise the distributions diverge quite a lot.

This is expected for the function, as grammatical constraints and lexical preferences are heavily language-dependent. Supersenses that only occur as function labels in a subset of languages are shaded in table 4. In particular, these are the GESTALT and POSSESSOR functions of possessives in German and English, the RECIPIENT function of datives in German, and the EXPERIENCER function in Mandarin Chinese. That SOURCE occurs more often as a function in German than in the other languages is because this includes almost all uses of the frequent adpositions *von* (including passive subjects) and *aus*.

On the other hand, the diverging scene role distributions may be due to topical differences between the annotated text samples. And even when expressing the same or similar meaning relations, languages vary in when and how they rely on using adpositions, as can be seen in the following examples:

(8) a.   Riesenschlagen verschlingen ihre Beute **in** einem Stück
         giant.snakes   swallow   their prey in one   piece
     b.   Boa constrictors swallow their prey **whole**

(9) a.   **Sein** Heimatplanet ist der Asteroid B-612
         his   home.planet is the Asteroid B-612
     b.   The planet he came **from** is Asteroid B-612

5 Impact on AI Applications—Three Case Studies

With our adapted annotation scheme, annotated corpus, and newly gained empirical insight, we lay the groundwork for applying SNACS supersenses to a variety of downstream tasks in German-speaking environments. To illustrate potential use cases, we discuss below three concrete scenarios involving human-robot interaction, machine translation (MT), and computer-assisted language education.[6]

5.1 Human-Robot Interaction

Adpositions and case are hard to interpret and act upon! Reasoning and communicating about objects in the real world comes natural to us as human beings, as we intuitively acquire embodied representations of scenarios, which deeply affect our language (see §2.3). In contrast, enabling a robot to engage in dialogue about its tasks, goals, and methods requires facilitating its symbolic and geometric reasoning about its environment first [22, *inter alia*]. Since a robot's executive and communicative faculties are not inherently intertwined in the way a human's are, its NLU representations

---

[6] These thought experiments are in their style and purpose inspired by Trott et al. [70].

and mechanisms need to be carefully engineered and tailored to its intended domain of application [45, 79]. It should be able to recover from fail states [34] and generate explanations about its own reasoning when prompted [67]. Such a framework, based on Abstract Meaning Representation [3], has recently been proposed by Bonial et al. [6].

As we have shown in §3 and §4, circumstantial, configurational, and participant relations are frequently expressed with adpositions and case in German. Here we look at a few examples where understanding this kind of language is crucial for successful task execution.

In this first exchange, the user requests a spatial transfer action from a mobile robot. She chooses to designate the RECIPIENT with a dative pronoun.

U: Fido, bring **mir**:RECIPIENT ein Glas Wasser!
     Fido, bring me a glass of water!

R: (Zu) wem:RECIPIENT soll ich das Wasser bringen?
           Wohin:GOAL
     To whom/where should I take the water?

It is reasonable for the robot to ask a grounding question about the RECIPIENT or GOAL of the transfer, provided it correctly identifies the scene-specific roles of the entities involved.

Consider now an alternative scenario with a stationary robot capable of producing artwork:

U: Dali, zeichne **mir**:BENEFICIARY ein Schaf!
     Dali, draw me a sheep!

*R: (Zu) wem:RECIPIENT soll ich ein Schaf zeichnen?
           Wohin:GOAL
     To whom/where am I supposed to draw a sheep?

R: Gerne zeichne ich ein Schaf für:BENEFICIARY dich!
     I'm happy to draw a sheep for you!

Here the same dative construction marks the BENEFICIARY: the product will benefit or be dedicated to the user. Given the user's communicative intent, selective preferences of the verb *zeichnen* 'to draw' corresponding to expectations about the evoked action of drawing, and of course the robot's physical abilities, a grounding question involving the RECIPIENT or GOAL of a transfer would be inappropriate. If, instead, the robot indicated that it successfully disambiguated the user's contextual role, the immersive experience would be greatly enhanced. However, this would constitute an extra requirement on the language interface design that goes beyond mere practical functionality since, in contrast to the previous example, the user is not part of this robot's actionable environment.

In order to avoid such misinterpretations, we recommend integrating supersense information into robots' language understanding (and generation) modules. There are many

	User input	System output	Plausible Scene Role(s)	Chosen
1.	Unter den Brettern war ein Schatz.	There was a treasure under the boards.	Locus*	Locus
2.	Unter ihnen war mein Schatz.	My darling was among them.	Whole*, Locus	Whole
3.	Unter den Rosen war ein Schatz.	Among the roses was a treasure.	Whole, Locus*	Whole
	Target:	Under	*(buried beneath the roses)*	
4.	Unter den Rosen war mein Schatz.	Under the roses was my sweetheart.	Whole*, Locus	Locus
	Target:	Among	*(the Little Prince is in love with a personified rose)*	
5.	Hier ist das Bild mal nachgezeichnet.	Here the picture is drawn.	Topic*, Path, Frequency	N/A
6.	Hier ist das Bild nachgezeichnet.	Here the picture is drawn.	Topic*, Path, Frequency	N/A
7.	Hier habe ich das Bild mal nachgezeichnet.	Here I have drawn the picture again.	Topic*, Path, Frequency	Frequency
8.	Hier habe ich das Bild nachgezeichnet.	Here I have traced the picture.	Topic*, Path, Frequency	Path
	Target:	Here is a copy of the drawing.	*(a new drawing modeled after the original)*	

Table 5: Examples of adpositional expressions translated by an automatic system, along with the scene roles of plausible, intended (asterisks), and chosen semantic readings.

methodological options for doing this. E.g., one could train and run an explicit supersense tagger in conjunction with a semantic parser [55], or directly optimize an end-to-end language model with respect to the latent posterior distribution of supersenses [42]. At least a subset of supersenses (e.g., Locus, Goal, Recipient) could also be explicitly mapped to a robot's extra-linguistic, i.e., visual and navigational representations.

5.2 Machine Translation

Adpositions are hard to translate! Hashemi and Hwa [23] found that prepositions were consistently among the largest sources of errors in texts automatically translated into English from German, Czech, Spanish, and French in the WMT 2012 shared task [9]. At the time, this problem was addressed, e.g., by Weller et al. [75, 74], who explicitly modeled semantic properties of prepositions and case in English-German translation. With the transition from statistical to neural MT (NMT), large pretrained transformer models have had great success in both quality and accessibility of fast automatic translation, though this by no means solves the problem entirely [54].

It is not hard to come up with examples of simple sentences whose adpositions confuse even industry-leading machine translation systems. A well-known freely available translation system[7] has no trouble disambiguating the different senses of *unter* ('under'/'among') in sentences 1 and 2 shown in table 5, but struggles with producing the intended interpretations of sentences 3 and 4.[8]

Which interpretations of *unter*—and thereby of the described scene—are available depends on the entity types of the subject (or conceptual trajector) *Schatz*, and the respective prepositional objects (or landmarks) *Brettern*, *ihnen*, *Rosen*. If trajector and landmark are of matching types (in terms of animacy, likelihood to be perceived as valuable, ...), the mereological sense (Whole) is more likely; and conversely, if certain expectations about size, shape, position etc. are (plausible to be) met given the description of the scene, the spatial sense (Locus) might be preferable. In our case, *Bretter* 'boards' are—rather unambiguously—inanimate objects with a certain prototypical size, shape, and purpose; *Rosen* 'roses' are plants to which culturally a high level of beauty or pleasantness is ascribed; *Schatz* can either be 'treasure', an inanimate but valuable object, or 'sweetheart', a highly-valued animate sentient entity; and *ihnen* 'them' is a pronoun that may assume the type of whatever entities it refers to.

An automatic system with access to massive amounts of raw text may well be able to pick up on such patterns purely based on having observed millions of similarly structured examples involving similarly distributed vocabulary, and so on. Yet, such a system clearly lacks any conceptual representation of the relations that hold between boards, treasures, roses, and sweethearts, neither in the real nor the metaphorical world—it chooses a statistically likely translation for the string *unter* given the other word strings and its model of word strings, without ever understanding what the strings mean.

This shortcoming becomes even more apparent in sentences 5–8. Sentence 5 is taken from the first chapter of our annotated Little Prince corpus, and slightly manipulated without dramatic changes to its meaning in sentences 6–8.[9] The system outputs vary wildly in how the verb particle *nach* is expressed (sentences 7 and 8), if at all (no part of the generated translations of sentences 5 and 6 corresponds to *nach*), even

---

[7] Google Translate

[8] The examples here are cherry-picked, and especially sentence 4 requires specific background knowledge, without which no automatic or human translator is expected to generate the intended output. However, by demonstrating that subtle differences in lexical meaning—even and especially that of adpositions—can have a dramatic impact on sentence meaning, we argue that translation systems that lack semantic awareness and nuance are inherently unreliable and prone to inconsistencies.

[9] *mal* is a modal particle establishing that the scene 'just happens to unfold this way'.

though we would expect near-identical translations for all four versions. The adposition *nach* itself is highly ambiguous, allowing both spatial DIRECTION and temporal INTERVAL function readings, each with multiple associated scene roles. As a verb particle in **nach***zeichnen* (lit. 'draw after') it can be translated as 'to trace' (emphasizing the PATH of the drawing motion), 'to draw again' (emphasizing the FREQUENCY of repeated drawing events), or 'to make a copy' (emphasizing that the content or TOPIC of a newly produced artwork resembles the existing one). These interpretations are all fairly similar, but their fine profiling distinctions are obvious to human readers.

Adding information about adposition supersenses to an automatic translation system, either as additional training data, or as an ontological resource at prediction time, could thus improve its adequacy, consistency, and fluency. As with the human-robot interaction scenario, the idea is to encourage the model to account for the adposition and its meaning in relation to the rest of the utterance, commit to an overall consistent and likely analysis, or at the very least recognize the potential for ambiguity and in turn suggest plausible alternative outputs.

5.3 Language Teaching and E-learning

Adpositions are hard to learn and teach! Adpositions and case are seldom the focus of automated or computer-assisted language teaching (or language instruction at all), and when they are, the learner is often just referred to lists or rules(-of-thumb), making it difficult to truly internalize their meaning [69, 49, 50, 78, 37]. This difficulty and, therefore, the need for high-quality pedagogy are highlighted in the academic literature, e.g., in DeHaven's dissertation on German two-way prepositions [15], Vyatkina [73] on German verb-preposition collocations, and Gradel [21] on the German case system.

To improve instructional methods, many researchers argue for cognitive and conceptual approaches where learners are encouraged to engage directly with the meaning of grammar rather than just the form [14, 21, 78].

We thus suggest equipping learning software with knowledge about the role/function distinction and conceptual semantic representations of adpositional expressions such as ours, in order to provide much more direct feedback to learners, for instance by retrieving concrete exemplars from a corpus or database.

6 Conclusion

We adapted the SNACS semantic annotation scheme to the analysis of German adposition and case semantics and piloted it in a novel corpus that will be released to the research community. Our framework is more comprehensive than previous approaches: alongside "classic" prepositions it includes several of their morphological variants as well as possessive and dative pronouns. By discussing examples and their SNACS analyses in the context of concrete NLP applications, we hope to have convinced the reader of the importance of treating these phenomena with proper care for the future of AI.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Arora A, Schneider N (2020) SNACS annotation of case markers and adpositions in Hindi. Presented at SIGTYP 2020: The Second Workshop on Computational Research in Linguistic Typology
2. Baker CF, Fillmore CJ, Lowe JB (1998) The berkeley framenet project. In: 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1, pp 86–90
3. Banarescu L, Bonial C, Cai S, Georgescu M, Griffitt K, Hermjakob U, Knight K, Koehn P, Palmer M, Schneider N (2013) Abstract Meaning Representation for sembanking. In: Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse, Association for Computational Linguistics, Sofia, Bulgaria, pp 178–186, URL https://www.aclweb.org/anthology/W13-2322
4. Bhattasali S, Fabre M, Luh W, Saied HA, Constant M, Pallier C, Brennan JR, Spreng RN, Hale J (2018) Localising memory retrieval and syntactic composition: an fMRI study of naturalistic language comprehension. Language, Cognition and Neuroscience URL https://doi.org/10.1080/23273798.2018.1518533
5. Blodgett A, Schneider N (2018) Semantic supersenses for English possessives. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), European Language Resources Association (ELRA), Miyazaki, Japan, URL https://www.aclweb.org/anthology/L18-1242
6. Bonial C, Donatelli L, Abrams M, Lukin S, Tratz S, Marge M, Artstein R, Traum D, Voss C (2020) Dialogue-AMR: Abstract Meaning Representation for dialogue. In: Proceedings of The 12th Language Resources and Evaluation Conference, pp 684–695
7. Brøndal V (1950) Théorie des prépositions: introduction à une sémantique rationnelle. Ejanr Munksgaard
8. Burchardt A, Erk K, Frank A, Kowalski A, Padó S, Pinkal M (2006) The SALSA corpus: a German corpus resource for lexical semantics. In: Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06), European Language Resources Association (ELRA), Genoa, Italy, URL http://www.lrec-conf.org/proceedings/lrec2006/pdf/339_pdf.pdf
9. Callison-Burch C, Koehn P, Monz C, Post M, Soricut R, Specia L (2012) Findings of the 2012 Workshop on Statistical Machine Translation. In: Proceedings of the Seventh Workshop on

Statistical Machine Translation, Association for Computational Linguistics, Montréal, Canada, pp 10–51, URL https://www.aclweb.org/anthology/W12-3102

10. Campe P (1997) Genitives and 'von-datives' in German: a case of free variation. Lexical and syntactical constructions and the construction of meaning pp 165–185

11. Cieschinger M (2016) The contraction of preposition and definite article in German—semantic and pragmatic constraints. PhD thesis, Universität Osnabrück

12. Cohen J (1960) A coefficient of agreement for nominal scales. Educational and psychological measurement 20(1):37–46

13. Crössmann H (1973) Präposition oder Konjunktion? Linguistische Studien 4:16–38

14. De Knop S (2015) Visualization and conceptual metaphor as tools for the teaching of abstract motion in German. Review of Cognitive Linguistics 13(1):167 – 190

15. DeHaven MR (2016) Input processing and the teaching of German two-way prepositions. PhD thesis, University of Kansas

16. Di Meola C (2004) The rise of the prepositional genitive in German——a grammaticalization phenomenon. Lingua 114(2):165–182

17. Durrell M, Brée D (1993) German temporal prepositions from an English perspective. The Semantics of Prepositions: From Mental Processing to Natural Language Processing 3:295

18. Erk K, Kowalski A, Padó S, Pinkal M (2003) Towards a resource for lexical semantics: A large German corpus with extensive semantic annotation. In: Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Sapporo, Japan, pp 537–544, DOI 10.3115/1075096.1075164, URL https://www.aclweb.org/anthology/P03-1068

19. Fillmore C (1967) The case for case. In: Texas Symposium on Language Universals

20. Gessler L, Wein S, Schneider N (2020) Supersense and sensibility: Proxy tasks for semantic annotation of prepositions. In: Proceedings of the 14th Linguistic Annotation Workshop, Association for Computational Linguistics, Barcelona, Spain, pp 117–126, URL https://www.aclweb.org/anthology/2020.law-1.11

21. Gradel V (2016) The acquisition of the German case system by foreign language learners through computer animations based on cognitive linguistics. Yearbook of the German Cognitive Linguistics Association 4(1):113–134

22. Gravot F, Haneda A, Okada K, Inaba M (2006) Cooking for humanoid robot, a task that needs symbolic and geometric reasonings. In: Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006., IEEE, pp 462–467

23. Hashemi HB, Hwa R (2014) A comparison of MT errors and ESL errors. In: LREC, pp 2696–2700

24. Haspelmath M (2003) The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. The new psychology of language 2(1976):1–30

25. Hofherr PC (2012) Preposition-determiner amalgams in German and French at the syntax-morphology interface. Comparative Germanic syntax: The state of the art Amsterdam: John Benjamins Publishing Company pp 99–132

26. Huddleston R, Pullum GK (eds) (2002) The Cambridge Grammar of the English Language. Cambridge University Press, Cambridge, UK

27. Hwang JD, Bhatia A, Han NR, O'Gorman T, Srikumar V, Schneider N (2017) Double trouble: The problem of construal in semantic annotation of adpositions. In: Proceedings of the 6th Joint Conference on Lexical and Computational Semantics (*SEM 2017), Association for Computational Linguistics, Vancouver, Canada, pp 178–188, DOI 10.18653/v1/S17-1022, URL https://www.aclweb.org/anthology/S17-1022

28. Hwang JD, Choe H, Han NR, Schneider N (2020) K-SNACS: Annotating Korean adposition semantics. In: Proceedings of the Second International Workshop on Designing Meaning Representations, Online

29. Hwang JD, Schneider N, Srikumar V (2020) Sprucing up supersenses: Untangling the semantic clusters of accompaniment and purpose. In: Proceedings of the 14th Linguistic Annotation Workshop, International Committee for Computational Linguistics, Barcelona, Spain (Online), pp 127–137, URL https://www.aclweb.org/anthology/2020.law-1.12

30. Jackendoff R (2002) Foundations of language: Brain, meaning, grammar, evolution. Oxford University Press, USA

31. Jespersen O (1963) The Philosophy of Grammar (Reprinted). George Allen & Unwin

32. Kipper K, Korhonen A, Ryant N, Palmer M (2008) A large-scale classification of English verbs. Language Resources and Evaluation 42(1):21–40, URL http://link.springer.com/article/10.1007/s10579-007-9048-2

33. Kiss T, Müller A, Roch C, Stadtfeld T, Börner K, Duzy M (2016) Ein Handbuch für die Bestimmung und Annotation von Präpositionsbedeutungen im Deutschen. Ruhr-Universität Bochum, Sprachwissenschaftliches Institut

34. Knepper RA, Tellex S, Li A, Roy N, Rus D (2015) Recovering from failure by asking for help. Autonomous Robots 39(3):347–362

35. Köper M, Schulte im Walde S (2014) A Rank-based Distance Measure to Detect Polysemy and to Determine Salient Vector-Space Features for German Prepositions. In: Proceedings of the 9th International Conference on Language Resources and Evaluation, Reykjavik, Iceland, pp 4459–4466

36. Koppel M, Ordan N (2011) Translationese and its dialects. In: Proc. of ACL-HLT, Portland, Oregon, USA, pp 1318–1326, URL https://www.aclweb.org/anthology/P11-1132

37. Kranzlein M, Manning E, Peng S, Wein S, Arora A, Schneider N (2020) PASTRIE: A corpus of prepositions annotated with supersense tags in Reddit international English. In: Proceedings of the 14th Linguistic Annotation Workshop, Association for Computational Linguistics, Barcelona, Spain, pp 105–116, URL https://www.aclweb.org/anthology/2020.law-1.10

38. Lakoff G, Johnson M (2008) Metaphors we live by. University of Chicago press

39. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33(1):159–174

40. Langacker RW, Tomasello M (1998) Conceptualization, symbolization, and grammar. The New Psychology of Language: Cognitive and Functional Approach to Language Structure pp 1–39

41. Levin B (1993) English verb classes and alternations: A preliminary investigation. University of Chicago press

42. Li XL, Rush A (2020) Posterior control of blackbox generation. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, pp 2731–2743, DOI 10.18653/v1/2020.acl-main.243, URL https://www.aclweb.org/anthology/2020.acl-main.243

43. Ljunggren KG (1951) Towards a definition of the concept of preposition. Studia linguistica 5(1):7–20

44. Malchukov AL, Narrog H (2008) Case polysemy. In: Malchukov AL, Spencer A (eds) The Oxford Handbook of Case, Oxford University Press, Oxford, UK, pp 518–534

45. Malmaud J, Wagner E, Chang N, Murphy K (2014) Cooking with semantics. In: Proceedings of the ACL 2014 Workshop on Semantic Parsing, Association for Computational Linguistics, Baltimore, MD, pp 33–38, DOI 10.3115/v1/W14-2407, URL https://www.aclweb.org/anthology/W14-2407

46. Manning CD, Surdeanu M, Bauer J, Finkel J, Bethard SJ, McClosky D (2014) The Stanford CoreNLP natural language processing toolkit. In: Association for Computational Linguistics (ACL) System Demonstrations, pp 55–60, URL http://www.aclweb.org/anthology/P/P14/P14-5010

47. Martins R (2012) Le Petit Prince in UNL. In: Calzolari N, Choukri K, Declerck T, Doğan MU, Maegaard B, Mariani J, Moreno A, Odijk J, Piperidis S (eds) Proc. of LREC, Istanbul, Turkey, pp 3201–3204, URL http://www.lrec-conf.org/proceedings/lrec2012/pdf/879_Paper.pdf

48. Masloch S (2020) Über "über" und dessen Entsprechungen im Englischen. Master's thesis, Ruhr-Universität Bochum

49. Mueller CM (2011) English learners' knowledge of prepositions: Collocational knowledge or knowledge based on meaning? System 39(4):480–490

50. Mueller CM (2012) Comparison of an integrative inductive approach, presentation-and-practice approach, and two hybrid approaches to instruction of English prepositions. Ph.D. dissertation, University of Maryland, College Park, Maryland, USA, URL https://drum.lib.umd.edu/handle/1903/12990

51. Müller A, Hülscher O, Roch C, Keßelmeier K, Stadtfeld T, Strunk J, Kiss T (2010) An annotation schema for preposition senses in German. In: Proceedings of the Fourth Linguistic Annotation Workshop, Association for Computational Linguistics, pp 177–181

52. Müller W (2004) Das Wörterbuch der deutschen Präpositionen. Zeitschrift für Germanistische Linguistik 32(3):406–438

53. Peng S, Liu Y, Zhu Y, Blodgett A, Zhao Y, Schneider N (2020) A Corpus of Adpositional Supersenses for Mandarin Chinese. In: Proceedings of The 12th Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, pp 5986–5994, URL https://www.aclweb.org/anthology/2020.lrec-1.733

54. Popović M (2017) Comparing language related issues for NMT and PBMT between German and English. The Prague Bulletin of Mathematical Linguistics 108(1):209–220

55. Prange J, Schneider N, Abend O (2019) Made for each other: Broad-coverage semantic structures meet preposition supersenses. In: Proc. of CoNLL

56. Puig Waldmüller ES (2008) Contracted preposition-determiner forms in German: Semantics and pragmatics. PhD thesis, Universitat Pompeu Fabra

57. Schmied J (1998) Differences and similarities of close cognates: English 'with' and German 'mit'. LANGUAGE AND COMPUTERS 24:255–276

58. Schneider N, Srikumar V, Hwang JD, Palmer M (2015) A hierarchy with, of, and for preposition supersenses. In: Proceedings of The 9th Linguistic Annotation Workshop, Association for Computational Linguistics, Denver, Colorado, USA, pp 112–123, DOI 10.3115/v1/W15-1612, URL https://www.aclweb.org/anthology/W15-1612

59. Schneider N, Hwang JD, Srikumar V, Green M, Suresh A, Conger K, O'Gorman T, Palmer M (2016) A corpus of preposition supersenses. In: Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-X 2016), Association for Computational Linguistics, Berlin, Germany, pp 99–109, DOI 10.18653/v1/W16-1712, URL https://www.aclweb.org/anthology/W16-1712

60. Schneider N, Hwang JD, Bhatia A, Srikumar V, Han NR, O'Gorman T, Moeller SR, Abend O, Shalev A, Blodgett A, Prange J (2018) Adposition and Case Supersenses v2: Guidelines for English. arXiv preprint arXiv:170402134v3 URL https://arxiv.org/abs/1704.02134v3

61. Schneider N, Hwang JD, Srikumar V, Prange J, Blodgett A, Moeller SR, Stern A, Bitan A, Abend O (2018) Comprehensive Supersense Disambiguation of English Prepositions and Possessives. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Melbourne, Australia, pp 185–196, DOI 10.18653/v1/P18-1018, URL https://www.aclweb.org/anthology/P18-1018

62. Schneider N, Hwang JD, Bhatia A, Srikumar V, Han NR, O'Gorman T, Moeller SR, Abend O, Shalev A, Blodgett A, Prange J (2020) Adposition and Case Supersenses v2.5: Guidelines for English. arXiv preprint arXiv:170402134v6 URL https://arxiv.org/abs/1704.02134v6

63. Scott A (2013) The genitive case in Dutch and German: A study of morphosyntactic change in codified languages. Brill

64. Shalev A, Hwang JD, Schneider N, Srikumar V, Abend O, Rappoport A (2019) Preparing SNACS for subjects and objects. In: Proceedings of the First International Workshop on Designing Meaning Representations, Association for Computational Linguistics, Florence, Italy, pp 141–147, DOI 10.18653/v1/W19-3316, URL https://www.aclweb.org/anthology/W19-3316

65. Smith MB (1995) Semantic motivation vs. arbitrariness in grammar: Toward a more general account of the DAT/ACC contrast with German two-way prepositions. Insights in Germanic linguistics I: Methodology in transition 1:293

66. Solstad T (2010) Post-nominal genitives and prepositional phrases in German: a uniform analysis. The syntax of nominalizations across languages and frameworks 23:219–251

67. Sridharan M, Meadows B (2019) Towards a theory of explanations for human–robot collaboration. KI-Künstliche Intelligenz 33(4):331–342

68. Stolz T (2007) *Harry Potter* meets *Le petit prince* – On the usefulness of parallel corpora in crosslinguistic investigations. STUF - Sprachtypologie und Universalienforschung 60(2):100–117

69. Takahaski G (1969) Perception of space and the function of certain English prepositions. Language learning 19(3-4):217–234

70. Trott S, Torrent TT, Chang N, Schneider N (2020) (Re)construing Meaning in NLP. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, pp 5170–5184, DOI 10.18653/v1/2020.acl-main.462, URL https://www.aclweb.org/anthology/2020.acl-main.462

71. Tyler A, Evans V (2003) The semantics of English prepositions: Spatial scenes, embodied meaning, and cognition. Cambridge University Press

72. Van Riemsdijk H (2007) Case in spatial adpositional phrases: The dative-accusative alternation in German. Pitar Mos: A building with a view Festschrift for Alexandra Cornilescu pp 1–23

73. Vyatkina N (2016) Data-driven learning for beginners: The case of German verb-preposition collocations. ReCALL 28(2):207—-226, DOI 10.1017/S0958344015000269

74. Weller M, Fraser A, Schulte im Walde S (2013) Using subcategorization knowledge to improve case prediction for translation to German. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Sofia, Bulgaria, pp 593–603, URL https://www.aclweb.org/anthology/P13-1058

75. Weller M, Schulte im Walde S, Fraser A (2014) Using Noun Class Information to model Selectional Preferences for Translating Prepositions in SMT. In: Proceedings of the 11th Conference of the Association for Machine Translation in the Americas, Vancouver, Canada, pp 275–287

76. Wescoat MT (2007) Preposition-determiner contractions: an analysis in optimality-theoretic lexical-functional grammar with lexical sharing. In: Butt M, King TH (eds) Proceedings of the LFG07 Conference, pp 439–459

77. Willems K (2011) The semantics of variable case marking (accusative/dative) after two-way prepositions in German locative constructions. towards a constructionist approach. Indogermanische Forschungen 116:324–366

78. Wong MHI, Zhao H, MacWhinney B (2018) A cognitive linguistics application for second language pedagogy: The English preposition tutor. Language Learning 68(2):438–468

79. Wächter M, Ovchinnikova E, Wittenbeck V, Kaiser P, Szedmak
S, Mustafa W, Kraft D, Krüger N, Piater J, Asfour T (2018) In-
tegrating multi-purpose natural language understanding, robot's
memory, and symbolic planning for task execution in humanoid
robots. Robotics and Autonomous Systems 99:148 – 165, DOI
https://doi.org/10.1016/j.robot.2017.10.012

80. Zwarts J (2005) The case of prepositions: Government and compo-
sitionality in German PPs. In: 21st Annual Meeting of the Israel
Association for Theoretical Linguistics, Haifa