Gunde Kurtz und Thomas Pötter

Generative semantics for prepositions, prefixes, and compounds

Abstract

This paper presents ideas for formally representing and working with prepositions as stand-alone constructs in connection with nouns and verbs, as prefixes, and as connecting constituents in compounds. As main starting points serve (a) Davis' proto(typical) roles for linking subject and complements of verbs with semantic roles in HPSG, (b) Pustejovsky's generative lexicon theory, and (c) Cyc's approach to hierarchical semantic relations. The objective is to reduce specification effort to a minimum while maximizing the possibilities to combine words, stems, and prefixes generatively: The semantics of prepositions in space, time, and logic as well as the principles of language-specific conceptualizations are put to use in hierarchical specifications. As much information as possible is stored in more general specifications. The issue is one of precisely constraining lexical entries and generatively explaining creative language use. Applications: (1) Machine translation – for each language prepositions are mapped only to and from the interlingual representation. (2) Disambiguation – the semantic structures of the preposition's arguments refine each other. (3) Selectional restrictions and the semantics of compounds – verbs, prepositions, and nouns carry syntactic as well as semantic information whose compatibility is checked by means of generative lexicon theory.

Zusammenfassung

Dieser Aufsatz präsentiert Ideen zur formalen Repräsentation und der Arbeit mit Präpositionen als alleinstehende Konstrukte in Verbindung mit Nomen und Verben, als Präfixe und als verbindende Konstituenten in Komposita. Als wesentliche Grundlagen dienen (a) Davis' Proto-Rollen, um Subjekt und Komplemente von Verben mit semantischen Rollen in HPSG zu verbinden, (b) Pustejovskys generative Lexikontheorie und (c) Cycs Ansatz zu hierarchischen semantischen Beziehungen. Ziel ist es, den Spezifikationsaufwand auf ein Minimum zu reduzieren bei gleichzeitiger Maximierung der Möglichkeiten, Wörter, Stämme und Präfixe generativ zu kombinieren: Die Semantik von Präpositionen in Raum, Zeit und Logik sowie die Prinzipien sprachspezifischer Konzeptualisierungen werden im Bereich hierarchischer Spezifikationen angewandt. So viel Information wie möglich wird in allgemeineren Spezifikationen gespeichert. Es geht darum, lexikalische Einträge präzise zu beschränken und kreativen Sprachgebrauch generativ zu erklären. Anwendungen: (1) Maschinelle Übersetzung – für jede Sprache müssen Präpositionen nur auf die Zwischensprache und zurück abgebildet werden. (2) Disambiguierung – die semantischen Strukturen der Argumente der Präposition werden gegenseitig verfeinert. (3) Selektionsrestriktionen und Kompositionssemantik – Verben, Präpositionen und No-

men tragen sowohl syntaktische wie auch semantische Information, deren Verträglichkeit mit Methoden der generativen Lexikontheorie überprüft wird.

Keywords: preposition, prefix, compound, proto-role, generative lexicon, hierarchical relation, constraint, default, HPSG, typed feature structure, TFS, sense extension.

1. Introduction

The starting point for this paper was (a) the observation that stems, prefixes, and affixes combine very productively and mostly regularly, (b) the wish to explain the analysis as well as the formation of words involving these phenomena to students of German as a foreign language, and (c) the wish to come up with a formal explanation of these phenomena in HPSG (Head-driven Phrase Structure Grammar) for disambiguation and interlingual machine translation.

If the meaning of the individual stems, prefixes, and affixes were known, then the regular parts of productive word fields could be completely specified just by giving a formal representation for the stem. A first example of such a word field is that containing the stem "deck": "bedecken" (to cover), "bedeckt" (covered), "aufdecken" (to reveal), "verdecken" (to conceal), "Zimmerdecke" (ceiling), "entdecken" (to discover), "Entdeckung" (discovery), "Deckblatt" (cover page).

Obviously, the word field containing the stem "spr?ch" contains similar sense-extensions and semantic combinations: "besprechen" (verb; to discuss, to talk about), "besprochen" (adjective; discussed), "versprechen" (to make a mistake in speaking; to promise), entsprechen (to correspond), "Besprechung" (conference), "Aussprache" (pronunciation), "Schnellsprecher" (fast speaker), "vielversprechend" (very promising).

Another phenomenon addressed is that of similar core meanings of prepositions in different domains: "im Haus" (in the house; locative), "im Mai" (in May; temporal), "in der Menge" (in the set; modal).

Prepositions are often used in an underspecified way. To fully understand the expression or to translate it in some language, a disambiguation is necessary: "Geruch im Schrank" (smell *in the interior* of the wardrobe), "Kleid im Schrank" (dress *on a coat hanger* of the wardrobe), "Riß im Schrank" (crack *in the (outside) wood* of the wardrobe).

Similar underspecification occurs in compounds that can be analyzed as being related through different prepositions: "Schrankfach" (compartment *in* the wardrobe) "Schrankdecke (ceiling *on top of* the wardrobe), "Schrankfuß" (foot *below* the wardrobe), "Schrankbemalung" (painting *on* the wardrobe), "Schrankschatten" (shadow *beneath/before/behind* the wardrobe). The explanation of these phenomena is based on the proposals of Davis (1996) and Pustejovsky (1995) for the formal specification of lexical semantics. The proto-roles given by Davis exhibit an appealing simplicity due to their conception as the minimal necessary set of roles for the interface between (linguistic) semantics and syntax. They can be considered a subset of the semantic roles used in Cyc (Lenat and Guha 1990,

Burns and Davis 1999) that can exist independent of particular linguistic applications. In a revised version, Davis (1999) associates the proto(typical) role attribute ACT(OR) with an entity *A* if *A* affects/changes/participates in/perceives/contains/possesses/is superior to an entity *B*. That entity *B is* then formalized as UND(ERGOER) only if it is affected/changed/moved or is an incremental theme. If entity *B is* conceived of/perceived by/resulting from/accompanied by *A*, *B*'s proto-role attribute is SOA (state of affairs). Further proto-role attributes – each representing only a single relation – are GR(OU)ND (path traversed), IMP-ON (forcefully impinged on), PART (included or part of/constitutes), INF (inferior to), POSSD (possessed by).

2. Assumptions, requirements, simplifications

In lexical semantics and HPSG syntax modelling several problems relevant to the solutions presented here haven't been clearly solved. For formal knowledge representation in the context of this paper, open issues are addressed as follows:

Lexical vs. world knowledge: The semantic information contained in the relations of this paper does not have to be lexical: It might as well stem from a semantic knowledge base, from a machine readable dictionary/encyclopaedia, a corpus or any inference chain involving these resources.

Advantages of an integration: All types of knowledge are easily accessible and available. That way many (syntactical and semantical) disambiguation problems can be solved in a single parsing step: Parsing should be faster and it is not necessary to make some information (in later steps) take precedence over other information (in earlier parsing steps).

Dangers: Knowledge about events and roles in the world should not contain linguistic information so that no semantic inference path becomes impossible due to a linguistic part of a precondition/constraint not being satisfied. It should be easily possible to generalize linguistic specifications to be valid for several languages. Idiosyncratic conceptualizations that apply to some individual languages or language groups should for the above reasons be kept separate from other linguistic or world knowledge.

Constraints as functions or hierarchical relations: Most HPSG researchers prefer to state knowledge in the form of attribute value matrices (AVMs) arranged into a type hierarchy and try to avoid functions in their representation. AVMs mostly require no additional specifications/implementations and can be processed in a very efficient way. On the other hand the number of such relations may explode or attributes in such AVMs may become more general in the context of strict typing. This is particularly true for the generalizations over space, time, and logic presented in this paper. When attributes are too general, it is no longer possible to state precise constraints. One solution may be not to inherit all constraints but to introduce a RESTR(ICTION) attribute which holds a set of constraints thus circumventing strict typing requirements. In this approach it is more difficult to set all coreferences correctly: One alternative are rule

systems – similar to those used in minimal recursion semantics (MRS) – that infer the proper constraints. Another complementary approach is to extend inheritance procedures to elements of weekly typed (constraint) sets.

Functional constraints with *n* arguments can be implemented as typed AVMs: Argument *i* with $1 \leq i \leq n$ can be made the value of the attribute ARG*i*. Alternatively, such functions might be implemented by other modules. E.g. in a robot, the vision subsystem could implement the function *on_straight_line_ordered* which checks if *n* points/objects are on a straight line in 3D space in the order of specification.

The concept of proto-roles in HPSG does basically correspond to the concept of functional constraints in generative lexicon theory. Using both in parallel is redundant. Proto-roles seem to be the best choice where the problems mentioned above don't exist. Stating all functional constraints from lexical semantics within proto-roles that were designed to represent syntactic information isn't elegant: It would be very hard to prevent the mixing of information from different quales in the qualia structure and of syntactic and semantic (possibly non-lexical) information.

The Cyc approach favours representing information in relations within a semantic hierarchy to exaggerated formal separations of different kinds of semantic knowledge. However, syntactic and semantic knowledge is kept clearly separate in Cyc. That seems practical.

Semantically plural complements: Some prepositions are restricted to either have several complements or a single complement that is semantically plural (e.g. "police"). For further processing coreferences to the CONTENT of all these requirements are needed as a list. So far, there seems to be no standard way to state this in HPSG. We use a construct "SemPlur-NP" that's supposed to solve this task and specify the requirements for its implementation below.

3. Generative semantics

The main idea is to abstract out spatial concepts that can be transferred into other domains like time and logic. A human mind can easily associate these with the symbols shown in the last column. One aim of this paper is to show how to transfer these notions to the computer. One exception among the semantic correspondences seems to be "um den Tisch" (around the table) which is inherently imprecise while "um 5 Uhr" is an exact specification. This has historical roots: Time used to be measured by sundials. Time specifications were always unprecise due to the unpreciseness of the shadow. Only in modern times this old saying has turned into a precise specification. However, the idea that temporal and modal/causal meanings are rooted in a spatial meaning isn't defeated.

The topmost seven prepositions ("in", "an", "auf", "über", "unter", "vor", and "hinter") in Figure 1 have static (dative case) and dynamic readings (accusative case). This classification in static and dynamic readings is not very relevant to the similarity of core meanings in spatial, temporal, and modal uses.

Figure 1 shows a scheme that has been used successfully in teaching foreigners the meaning of German prepositions.

	Spatial	Temporal	Logical (modal/causal)	Symbol
in (in)	im/ins Zimmer	im Mai (in 2 Jahren)	einsammeln in Eile; in der Menge	
an (on)	an der/die Wand	am Sonntag	anlernen denken an	
auf (on)	auf dem/den Stuhl	auf Jahre hinaus	auflegen auf deutsch	
über (over, above)	über der/die Tür	über Nacht	überlegen sprechen über	
unter (under)	unter dem/den Tisch	unter der Woche	unterlassen unter Druck	
vor (in front of)	vor der/die Uhr	vor 2 Jahren	vordenken vor Neid	
hinter (behind)	hinter dem/den Busch	———	hinterlassen hinter etwas kommen	
nach (after, behind)	nach der ersten Person	nach 2 Tagen	nachholen trachten nach	
zu (to, towards)	zur Post	zur Zeit	zusehen zu Fuß	
bis (till)	bis Paris	bis Mittag bisweilen	———	
bei (near, with)	beim Chef	bei Tag	beitreten bei Frost	
gegen (against)	gegen die Wand	gegen Mittag	gegenzeichnen stimmen gegen	
zwischen (between)	zwischen den Stühlen	zwischen 10 und 12 Uhr	zwischenfinanzieren Vergleich zwischen	
um (at)	um den Tisch	um 5 Uhr	umarbeiten es geht um	
ab (from .. on)	ab Mainz	ab Montag	abraten ab Leutnant ist man..	trennen
von (from)	von zuhause	vom Vortag	sprechen von vom Blatt	Ausgangs-punkt
aus (of, out of)	aus dem Haus	aus den Achtzigern	ausarbeiten aus Holz; aus Mitleid	Menge

Figure 1: German prepositions with similar semantics in space, time, and logic.

For computers a type hierarchy with appropriate features can fulfill a similar task. Figure 2 represents such an attempt. For individual languages, this fairly general interlingual hierarchy can be further refined.

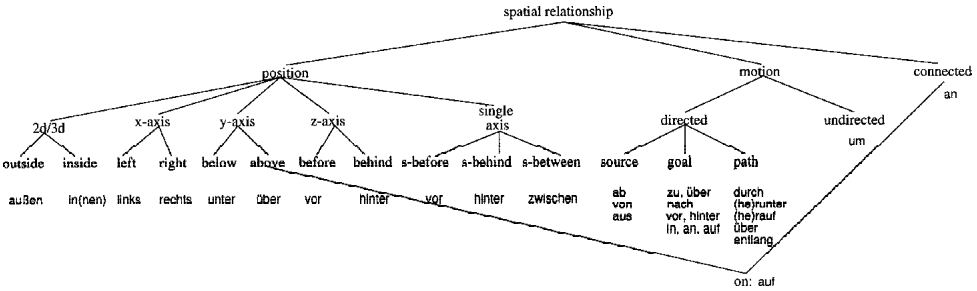


Figure 2: Interlingual semantic type hierarchy for spatial concepts.

All leaf nodes in Roman font with English names contain an attribute, which has the same name as the type. Below the English type names are some examples for German prepositions belonging to those types. Some prepositions like “in, an, auf, über” have several supertypes or require further disambiguation: Are they directed or not? “nach” is used with countries and cities as objects while “zu” is chosen for persons as objects. Of course, multiple inheritance is allowed, as illustrated for “auf” which inherits both from *connected* and *above.* For each language one might have to refine – but not restructure – this hierarchy in a slightly different way.

Here is a short summary of the most important properties of these spatial relation classes in English and German - based on Trujillo (1995: 171ff.):

Relation *goal* describes a motion endpoint - which is naturally *bounded*, see the table on features of prepositions below. One may infer for all sentences containing bounded motion verbs V like “walk” that “Subject V-ed to X” => “Subject was in/on/at X at time t_2” and “Subject was not in/on/at *X* for all times $t_1 < t_2$”.

(1) He drove through the Brenner tunnel to Italy.

Paths but not directions can be modified by goals as in (1). One can infer that for all t with $t_1 < t < t_2$: “Subject V-ed through/across/along *X*” => “Subject was in/on/at *X*”. The preposition “through” seems to always select “on” in this type of conclusion.

Directed prepositions can act neither as argument nor as functor of other relations (not even the goal relation). Neither can one infer that “Subject was in/on/at *X*” from a phrase containing a directed preposition. They state the orientation but not the endpoint of a motion:

(2) He was on the way to the top.

Prepositions which are subtypes of *position* can – at least in German – be the endpoints of motions. This case can be handled by making these “positions” inherit some motion properties or by using lexical rules.

If semantic hierarchies are available, translation involves the following steps: Find the most specific semantic type to which the preposition belongs in the source language. Look up if the verbalization in the target language is uniquely determined for that type. If not, disambiguate the preposition in the source language so that the semantic constraints for one unique subtype (representing a correct translation) in the target language hold.

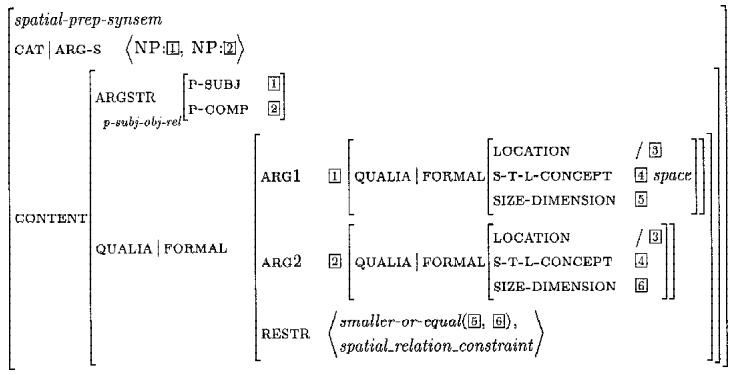
If it is known that in a specific language all subconcepts of a term are conceptualized as *n*-dimensional spaces, this fact should be stored in the most general concept for which this is true so that it can be inherited by many other subconcepts.

In English “floor” is visualized as two-dimensional entity (just the part one walks on) just as “Flur” in German. The German words “Stock” and “Stockwerk” are associated with a three-dimensional concept. Thus one walks “on floor 5” and also the German “Flur” takes “auf” but “Stock” takes “in”.

However, it may also be necessary to establish completely new hierarchies of prepositions for temporal, modal, or causal concepts in some languages. This can be done by introducing at the top level or a very high level concepts like *time* or *cause.*

For the typical spatial preposition with a single complement lexical entries inherit the information of the structure in (3).

(3)



P-SUBJ(ECT) and P-COMP(LEMENT) get assigned to the semantic content of the prepositional phrase's (PP's) subject and complement, respectively. This formalization favours the more general attribute P-COMP for prepositions instead of the several PART, INF, GRND or SOA which might differ for corresponding prepositions across natural languages. Such a change of a proto-role attribute makes it impossible to specify a useful interlingual common supertype relation.

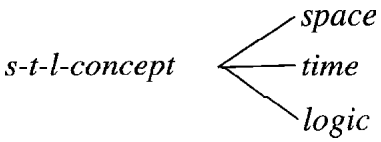


Figure 3: Subtypes of s-t-l-concept.

The FORMAL quale (following Pustejovsky's qualia theory) generally contains distinguishing features within a larger domain. In the path QUALIA | FORMAL it contains an attribute S-T-L-CONCEPT which can have the three subtypes *space, time,* and *logic* as shown in Figure 3.

Specification (3) contains the constraint that the LOCATION entry in subject and complement must be identical (coreference 3). This means that their values are unified and that parsing fails if the unification of these two LOCATION values is not possible. As described below, that unification performs the disambiguation desired. However, these locative specifications are only defaults – indicated by the slash "/": A wood-worm is usually in the outside wood of the wardrobe but it may also fall onto a piece of clothing.

This specification does also account for the (un)grammaticality of the following examples and helps in disambiguating some examples:

(4) The pencil (spatial) is in the drawer (spatial).

(5) *Look at the may (temporal) in the room (spatial).

(6) *The chair (spatial) in the table (modal).

(7) The chair (spatial) in the set (disambiguation: movie set [spatial] instead of a mathematical set [logical]).

The constraint that the prepositional subject (P-SUBJ role) must be smaller or equal in its size dimension (microcosm, real world, macrocosm) than the prepositional complement (P-Comp role) – coreferences 5 and 6 – helps to disambiguate sentences like:

(8) She watches the star *on* a hill *with* a telescope.

In this sentence the attachment of both prepositions is syntactically ambiguous: Both could refer to "she" or "the star". Since a "star" belongs to macrocosm while both complements of the two prepositions belong to the real world, the *smaller-or-equal* constraint above rules out that any of the prepositions modifies "the star".

One example for a preposition which has similar semantics in most of its interpretations is "in". What distinguishes all interpretations of "in" from general spatial prepositions (3) in this approach is shown in (9). It represents an approach to model the meaning of "in" as pictured in the first line of Figure 1.

(9)
$$\left[\begin{array}{l} \textit{in-type-synsem} \\ \text{CONTENT} \mid \text{QUALIA} \mid \text{FORMAL} \left[\begin{array}{ll} \text{ARG1} & \boxed{1}\left[\text{QUALIA} \mid \text{FORMAL} \mid \text{S-T-L-CONCEPT} \quad \boxed{3}\right] \\ \text{ARG2} & \boxed{2}\left[\text{QUALIA} \mid \text{FORMAL} \mid \text{S-T-L-CONCEPT} \quad \boxed{3}\right] \\ \text{RESTR} & \left\langle \textit{smaller-or-equal}(\boxed{5}, \boxed{6}), \right. \\ & \left. \textit{contained\_in}(\underset{\textit{s-t-l-concept}}{\boxed{3}}, \boxed{1}, \boxed{2}) \right\rangle \end{array}\right] \end{array}\right]$$

In this case it is specified that the S-T-L-CONCEPT *(space-time-logic-concept)* must be the same for both arguments and that the P-SUBJ must be semantically *contained in* the P-COMP. The core meaning of "in" is the spatial sense (10).

The spatial reading of "in" comes again with several complications. Even this pretty specific meaning needs disambiguation under many circumstances, especially when it has to be translated into other languages which distinguish senses in a more subtle way. Some examples are given in (11)-(17), partly based on Herweg (1989:111):

(11) Die Socken im Schrank (im Fach/in den Schubladen des Schrankes; the socks in the compartment/drawer of the wardrobe).

(10)
$$\left[\begin{array}{l} \textit{in-spatial-synsem} \\ \text{CAT} \mid \text{ARG-S} \quad \langle \text{NP:}\boxed{1}, \text{NP:}\boxed{2} \rangle \\ \text{CONTENT} \left[\begin{array}{ll} \text{ARGSTR} & \left[\begin{array}{ll} \text{P-SUBJ} \mid \text{CONCEPT} & \underset{\textit{spatial-entity}}{\boxed{1}} \\ \text{P-COMP} \mid \text{CONCEPT} & \underset{\textit{8d-9d-spatial-receptacle}}{\boxed{2}} \end{array}\right] \\ \text{QUALIA} \mid \text{FORMAL} & \left[\begin{array}{ll} \text{ARG1} & \boxed{1}\left[\text{QUALIA} \mid \text{FORMAL} \left[\begin{array}{ll} \text{S-T-L-CONCEPT} & \boxed{3}\ \textit{space} \\ \text{LOCATION} & / \boxed{4} \end{array}\right]\right] \\ \text{ARG2} & \boxed{2}\left[\text{QUALIA} \mid \text{FORMAL} \left[\begin{array}{ll} \text{S-T-L-CONCEPT} & \boxed{3} \\ \text{LOCATION} & / \boxed{4} \end{array}\right]\right] \\ \text{RESTR} & \textit{contained\_in}(\underset{\textit{spatial-concept}}{\boxed{3}}, \boxed{1}, \boxed{2}) \end{array}\right] \end{array}\right] \end{array}\right]$$

(12) Die Schuhe im Schrank (am Boden des Schrankes; the shoes at the bottom of the wardrobe).

(13) Das Kleid im Schrank (auf einem Kleiderhaken auf der Stange im Schrank; the dress on a coat hanger on the clothes bar in the wardrobe).

(14) Der Geruch im Schrank (im Innenraum des Schrankes; the odour inside the wardrobe).

(15) Der Riß im Schrank (in der Verkleidung des Schrankes; the crack *on* the outside wood of the wardrobe).

(16) Der Holzwurm im Schrank (im Holz des Schrankes; the wood-worm in the wood of the wardrobe).

(17) Die Motte im Schrank (in einem Kleidungsstück im Schrank; the moth in an article of clothing within the wardrobe).

Such finer distinctions are necessary to disambiguate for preposition attachment and semantic refinement – especially in the context of machine translation: In French, (15) should be translated with “sur” or “de” instead of “dans”. (16) requires the more precise wording “dans le bois” (in the wood) because the worm would otherwise be conceived as being inside the wardrobe (like shoes). In Hindi, (14) requires a different construction meaning “originates from”. Hindi differentiates the subjects of *in-spatial* into entities that are movable (preposition “ke ander”) and non-movable (preposition “me”).

Mongolian differentiates between “dotor” for “inside an interior space” (clothes, shoes in the wardrobe) and “dech” for “in the material” (crack, odour, wood-worm).

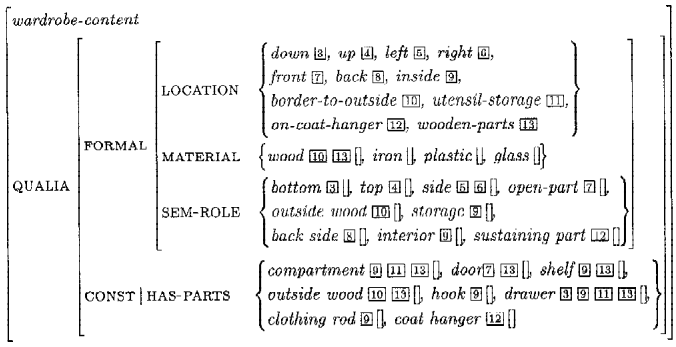
In Chinese the translation for the crack sentence (15) differs from the others in that it requires an explicit introduction of a “surface” notion.

In Russian and Bulgarian the case assignment differs for the various objects in the wardrobe. This shows that it is necessary to disambiguate the meaning of prepositions because there is either no standard translation of “in” or because that standard translation is not acceptable in some cases.

(18) shows how a part of the lexical entry for *wardrobe* stored in SYNSEM | LOCAL | CONTENT might look.

This information is perfectly suited to perform the disambiguation tasks required to interpret sentences (11)-(13):

(18)



For *socks* (11) one might store in the FORMAL | LOCATION quale the value *utensil-storage*, which might in turn be inherited from a class *utensil* or *household-item*. Through the constraint imposed by *in-spatial-synsem* in (10), the slot SYNSEM | LOCAL | CONTENT | QUALIA | FORMAL | LOCATION has to be equal in the formal arguments of the subject (socks) and the complement (wardrobe). Thus the LOCATION entry in *socks* selects the corresponding LOCATION entry in *wardrobe: utensil-storage*. This is coindexed with *compartment* and *drawer* – the inference needed to correctly analyze and translate the sentence into some languages.

Alternatively, the information about *socks* (11) that they are normally stored in a *compartment* or *drawer* might be stored in SYNSEM | LOCAL | CONTENT | QUALIA | FORMAL | SEM-ROLE *compartment*. This type of information usually isn't lexical: It would rather be omitted in favour of more prototypical information about socks. However, one can infer this from corpus processing where a sentence like "Sally put her socks back into the drawer", or "put the socks in the top drawer" or a wardrobe description containing the sentence "it has two socks drawers" occurs. If the system has a good "encyclopaedic" or "world knowledge" base (e.g. Cyc), it might be able to infer that a drawer is the only logical place for socks, since placing them elsewhere in the wardrobe wouldn't make sense: They would fall off the coat hanger, roll around at the bottom or in a compartment. To make the lexical specification a little less redundant by using abstraction, *shoe* (12) could inherit the location *down* as default (identified by "/") from *physical object*: SYNSEM | LOCAL | CONTENT | FORMAL | LOCATION*/down:* Because of the gravitational force physical objects are normally located on other objects or on the location which is a *down* location. Consequently, "shoe" selects the *down* location when occurring together with "in". Since *down* is coindexed with *bottom* in the semantic specification of *wardrobe*, the correct reading for (12) – being located at the bottom – can be inferred.

For *dress* (13) – *similar* to the "socks" – one might be able to infer from sentences in a corpus or from the knowledge base that the LOCATION value could be "*on-coat-hanger*". As for all spatial prepositions, the LOCATION attribute of subject and complement must agree. In this case the complement would have the more precise LOCATION value which helps in the selection process: *Coat hanger* could again have a LOCATION value of "*on-wardrobe-bar*". Finally, the LOCATION value of *wardrobe-bar* could be "*in-wardrobe*".

Alternatively, it might be stored or inferred for *dress* (13) that it's typical location is *on-coat-hanger*. Within the corresponding slot of *wardrobe*, this would select *sustaining-part* which is coindexed with *coat-hanger*. As in the first variant, inference could continue with the information stored in the lexical entry for *coat-hanger*.

An *odour* (14) is a kind of gas and as such mixes with the entire inside volume of a closed body. Thus LOCATION is equal to *inside* for *odour*. When unified with the data for *wardrobe*, this attribute turns out to be coindexed with *interior*, the disambiguated knowledge required for machine translation in some target languages.

Cracks (15) are typically on the *border to the outside* of an object, which could automatically be inferred from the WordNet definition for "scratch/crack": "a depression

scratched or carved into a surface". Unified with the lexical data for *wardrobe* this yields the coindexed SEM-ROLE *outside wood* made of MATERIAL *wood.*

The LOCATION slot for *wood-worm* as in (16) might contain *wooden-parts.* This information can in principle be inferred from the fact that *(a)* worms are generally contained in objects (slot LOCATION *material)* and *(b)* that the "wood" part of "wood-worm" stands for the material in which the worm is contained. The methods of compound analysis that achieve this are explained in Section 4.

From the coindexation with the LOCATION slot in *wardrobe,* it can be inferred that the MATERIAL *is wood* and the possible parts are: *compartment, door, outside wood, shelf, drawer.*

For interpreting sentence (17) correctly, it would be sufficient to have the lexical entry of *moth* hold the locative specification *article-of-clothing.* The lexical information of *article-of-clothing* might have as subtypes: dress, pants, shirt, underwear, socks, etc. As illustrated above, these types could again hold information about where they are typically stored. From that inference chain one might also conclude that the moth might be in the compartment/drawer containing the socks or near the clothing rod sustaining the dress.

3.1 Prepositions as prefixes

Many prefixes are very closely related to prepositions. "ein-" is derived from "in", and many other prefixes like "aus-", "über-", "durch-", "unter-", "um-", etc. are orthographically as well as semantically equivalent to prepositions. The possible subjects and objects of the unprefixed main verb determine the type of possible prefixes that a verb can take if the type of sense derivation is regular.

Examples of verbs that can be prefixed with "ein" (in) are: Spatial meaning: einlassen, einsehen, einstecken, einbrechen, einbrocken, eindringen, einfügen, einfließen, einsammeln.

Abstract meaning: einarbeiten, einbeziehen, einbilden; generally with abstract meaning: "ein" + adjective: "einfeuchten", "eindeutschen", "eindicken".

(19) Er bricht ein / Er hat eingebrochen (He breaks [into the house]; He has broken [into the house]).

(20) Man sollte das Altglas in den Glascontainer einwerfen (One should throw the waste glass into the glass container).

One way to formalize this meaning is shown in Figure 4 for the case of an intransitive base verb (cf. (19)) and in Figure 5 for a transitive base verb (cf. (20)).

The orthography/phonology of the prefixed verb is made up of the prefix "ein" concatenated with the orthography/phonology of the morphological base MORPH-B. Coreference 2 states that the semantic content of the subject, equal to the NP[nom] in the ARG-S slot takes the semantic UNDERGOER rule. The specification also states that the new process denoted by the "ein"-verb is directed and has a goal which is not specified any further.

In fact there are a few examples which have no precise ENDP(OIN)T: (21) or are not DIRECTED: (22), (23).

(21) Er steigt ins Haus ein.

(22) Sie steckt den Schlüssel ein.

(23) Sie fügte das Wort in Zeile 4 ein.

These must have an extra formal semantic specification with less information in the FORMAL quale. There is no phrasal constituent predictable whose semantic content should be equal to the points described by coreference 4 since it is an intransitive verb. This is different for transitive verbs.

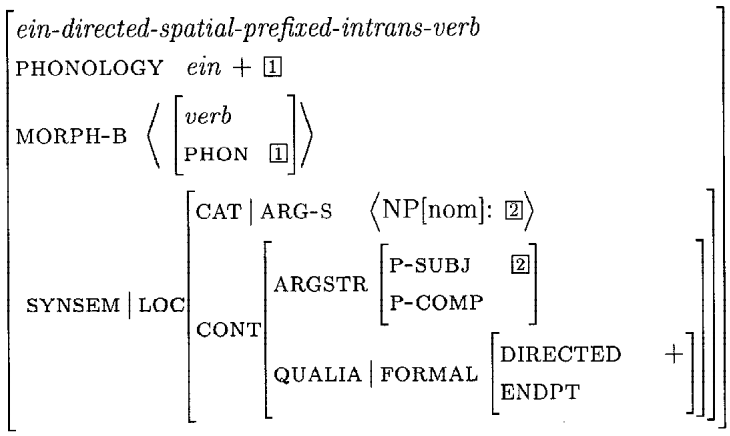


Figure 4: Formal semantics of the prefix "ein" selecting for an intransitive verb.

As shown in Figure 5, some additional statements can be made about transitive verbs that are directed and have endpoints: The semantic content of their complement in accusative case is equal to the ENDP(OIN)T feature in the FORMAL quale. This complement, as stated by coreference 3, is also equal to the complement of the base verb, taking the semantic GRND role in its ARGSTR.

MORPH-B is the morphological base; here it is the word without the prefix.

The next preposition to be considered is "zwischen" (between, among), previously also discussed in Habel (1989) and Trujillo (1995). Some example sentences are:

(24) Der Pfeffer steht zwischen Salz und Majoran. (The pepper is located between salt and marjoram; spatial usage)

(25) Könnten wir uns zwischen Montag und Mittwoch treffen? (Could we meet between Monday and Wednesday? Temporal usage)

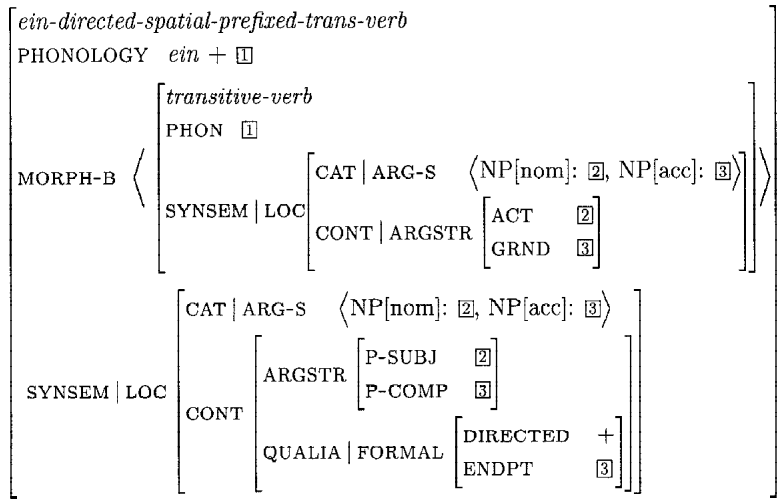


Figure 5: Formal semantics of the prefix "ein" selecting for a transitive verb.

(26) Kalter Krieg ist ein Zustand zwischen Krieg und Frieden. (Cold war is a state between war and peace; modal usage)

(27) *Der Soldat steht zwischen Krieg und Stuhl. (The soldier stands between war and chair.)

(28) *Zwischen Montag und Tisch werden wir uns treffen. (We'll meet between Monday and table.)

The formalization is shown in Figure 6.

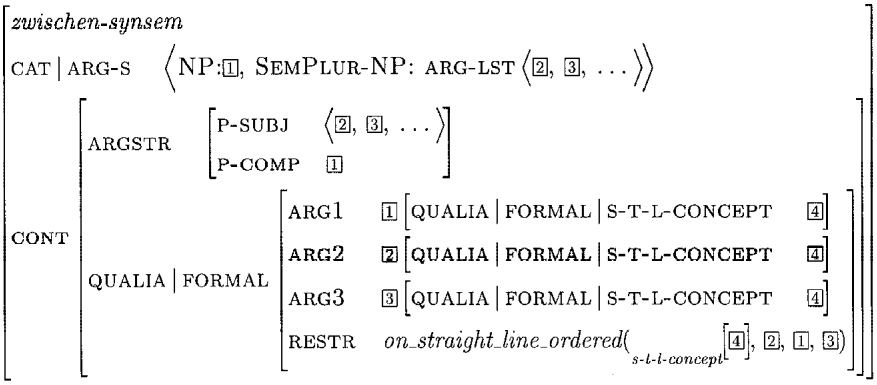


Figure 6: Formal semantics of the preposition "zwischen" (between) with two complements.

In Figure 6 the semantic P-SUBJ(ECT) is coindexed with the content of the subject, the first element in the ARG-S list – in a style similar to that of Tony Davis. Both complements are realized as ARG1 and ARG2 in the formal quale. S-T-L-CONCEPT stands for "space-time-logic-

concept". The coreference 4 assigned as its value states that this concept must be equal in both complements where it is stored in the formal quale: If one complement stands for a temporal entity, then all other complements must also represent a temporal entity. Such a constraint does also help in disambiguation tasks. This constraint accounts for the ungrammaticality of sentences (27)-(28). The RESTR(ICTION) in the formal quale states that the semantic content of all NP complements and the subject (coreference 1) must denote objects (in either space, time or logic – as indicated by coreference 4) which are conceptually on a straight line and must occur in the order given. It may be the case that no second complement (coreference 3) or even more than two complements exist. These cases are supposed to be handled by the SemPlur-NP semantically plural noun phrase).

SemPlur-NP must have at least the following properties:

1. It must represent an enumeration separated by commata or "und" (and).
1 The semantic roles of the potentially unbounded number of syntactic arguments must be extracted into the argument list (ARG-LST) of semantic arguments. This list is used to formulate further semantic constraints originating from the preposition (e.g. "zwischen").
3. When an "oder" (or) occurs in an enumeration, it must be treated semantically as if the preposition occurred several times: (30).
4. It may allow a single complement which must be semantically plural: (29).
5. Nominalizations and figurative readings must be handled correctly: (31).

(29) zwischen dem Gemüse (among the vegetables)

(30) zwischen dem Gemüse *oder* den Früchten (among the vegetables or the fruits) ⇔ zwischen dem Gemüse *oder* zwischen den Früchten (among the vegetables or among the fruits)

(31) zwischen Gut und Schlecht gibt es viele Zwischenstufen (there are many intermediate steps between good and bad)

The most promising way to implement this seems to be in a hierarchical fashion with several subtypes to handle each of the cases described above. One way to accomplish an unbounded number of complements is to specify the SemPlur-NP or a subtype of it again as element in the COMPS list. The implementation must be consistent with that of "and" and "or".

"Zwischen" can be translated either as "among" or "between" in English, as examples (32)-(37) show:

(32) Fred sits *among* students. (Fred is a student)

(33) Fred sits *between* students. (Fred is not necessarily a student)

(34) There is a shell *among/between* the stones.

(35) *Among* the universities Stanford is famous for HPSG.

(36) Our car is the one *between* the red and the black car.

(37) *Between* source and ocean the river travels for 1000 km.

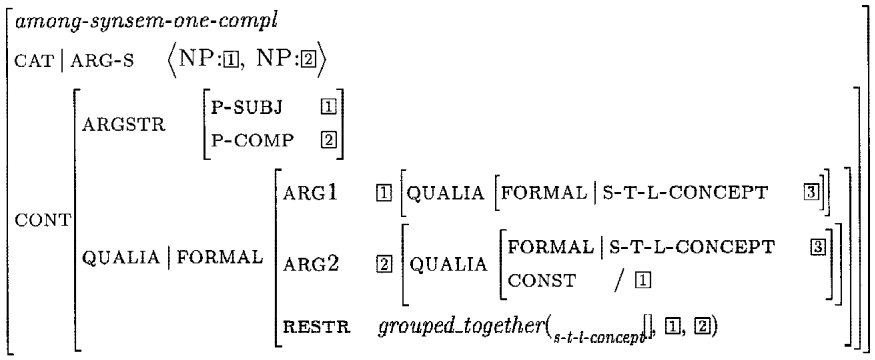


Figure 7: Formal semantics of the preposition "among" with one complement.

The analysis of these cases is derived from the findings presented in Trujillo (1995) in a different formalization.

As sentences (32), (35) illustrate, "among" is preferred when the FORMAL object (ARG1, usually the subject) is *part of* the P-SUBJ(ECT) object (ARG2, usually the complement of "among"). This *part of* relation is stated by coreference 3 as a default. The reason is that "among" can also be used under other circumstances when "between" is not favoured. Checking that a *part-of* relation holds between the complements can be done with the help of the CONST(ITUTIVE) quale; (Trujillo 1995:225) is similar. Again, the FORMAL quale is made equal to the semantics of the first element in the ARG-S list and ARG2's FORMAL slot by coreference 1. Coreference 3 specifies that the S(PACE)-T(IME)-L(OGIC)-CONCEPT has to be identical for both arguments.

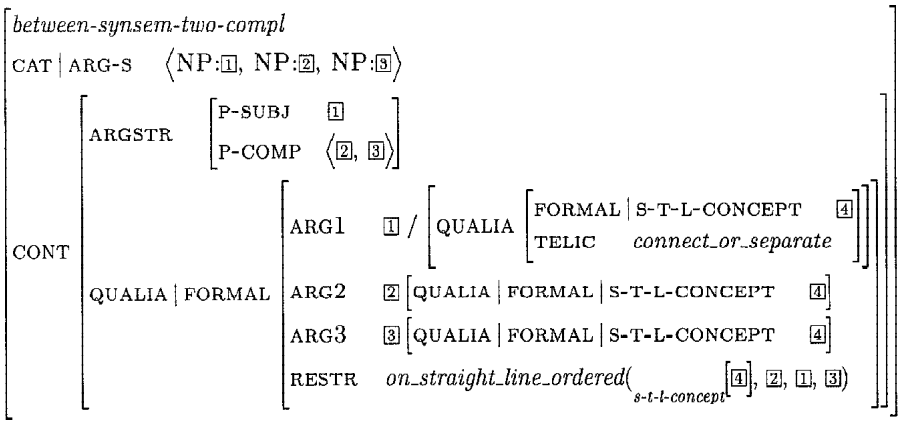


Figure 8: Formal semantics of the preposition "between" with two complements.

Sentences (33), (36), (37) show that "between" is preferred with arguments that are connecting or separating nouns, as Trujillo notes. In Figure 8 this is specified in the FORMAL quale for ARG1 (normally the subject of the sentence): It's TELIC quale is by default of the conjunctive type *connect_or_separate* – the minimal superconcept of *connect* and *separate*. Alternatively,

two separate versions of “between” could be stored in the lexicon (Trujillo’s approach) or a real conjunction could be used. Several researchers – among them Ann Copestake (1992, 1993) – state that conjunctions should generally be avoided because of the high expenses to implement and execute reasoning operations on them.

There exists a version of “between” with just one complement. This is completely analogous to the corresponding entry for “zwischen”, given in Figure 6 above.

A slight complication for this analysis is presented below:

(38) Mit Toyota an die Spitze der Pannenstatistik. (With Toyota to the top of breakdown statistics)

“Spitze” (top) in sentence (38) is a (figurative) sense extension of the spatial concept “top”. If “mit” (with) had the same constraint as “zwischen” (between) that both of its arguments have to be of the same *space-time-logic-concept*, then this sentence would not be accepted: “Spitze” (top) is a spatial concept while “Pannenstatistik” (breakdown statistics) belongs to “logic” in that classification.

In this case of statistics, the “top” is “the first line”. In fact it may be necessary to translate this as “first line of the breakdown statistics” in some languages. This case can be solved in either one of the following ways:

1. Type coercion as proposed by Pustejovsky: In the context of statistics or lists, potentially listed entities are coerced to modal concepts as part of a sense extension rule. In this case “top” and “Toyota” would be converted.
2. Lexical rules to carry out the conversion between the different concepts (only) under very constrained circumstances; In this case the same conversion as under 1 could be done;
3. Idiosyncratic specification in the lexicon (i.e. multiple lexical entries or semantic concepts for “Spitze”);
4. A formal semantic specification for “mit” (with) or for a subclass of “mit” which captures this case.

4. A generative analysis of German compounds

This idea can be carried even further towards analyzing compounds, see Table 1. Instead of analyzing one part of the complement as the head and others as modifiers, it is possible to treat as head the preposition which is usually omitted. The set of candidate prepositions can be narrowed down by requiring them to be unifiable with the word constituents that it is supposed to combine. With all the constraints imposed on the prepositions above, this set will be very small. That implicit preposition will very likely appear in the translation of the whole compound.

The Minimal Recursion Semantics approach (Copestake et al. 1997) lends itself for formalizing “Flurnachbar” (floor neighbour) (39).

Modifier concept	Relation	Example
Material	made of "aus"	Pelzmütze (cap *made of* fur)
Adjective	has property	Bleichgesicht (face *that is* pale)
Feeling	because of "vor"	Freudentränen (tears *of* joy)
Animate	for "für"	Mausefalle (trap *for* mice)
Society	belonging to (genitive)	Vereinsvorsitzender (chairman *of the* society)
2 nouns of same category	with "und"	Strichpunkt (semicolon; comma *with a* colon)
River/Ice/Road	on "auf"	Donauschiffahrt (shipping *on* the Danube)
Temporal expression	at the time of/"in"	Aprilwetter (weather *in* april)
Location/sea	"in"	Atlantikfischerei (fishing *in* the atlantic)
Location	"auf"	Schrankdecke (ceiling *on top of* the wardrobe)
Location	"in"	Schrankfach (compartment *in* the wardrobe)
Location	"unter"	Schrankfuß (foot *below* the wardrobe)
Location	"auf"	Schrankbemalung (painting *on* the wardrobe)
Location	"neben/um"	Schrankschatten (shadow *beneath/before/behind* the wardrobe)

Table 1: Common types of nominal compounds in German.

LANG is the superclass of all language names containing an attribute of the same name. It is supposed to unify with all language names like ENGLISH and GERMAN. If that's not possible, some alternative modelling possibilities are (a) distributed disjunctions, (b) the value of CONCEPT is considered a set or better a choice system (aka "choose-one-of"; known from systemic functional linguistics), (c) DATR global inheritance / evaluable paths, (d) hierarchical relations (similar to proto-roles) that contain both LANG(UAGE) and CONCEPT(UALIZATION) as attributes with the values shown in (39), (e) a finite state automaton (FSA), e.g. containing the language names on the upper band (upper label of the graph edge) and the associated spatial concepts on the lower band (lower label of the graph edge).

The MRS attributes used in (39) are: HANDEL – A unique id for each constituent. Named in honour of the musician Georg Friedrich Händel. INDEX – Used for saving lambda variables. INST – Instance of; used for semantic relations between quantor – subject – predicate – complements – adjuncts. LISZT – A list of predicates which are valid at the same time. Named in honour of the musician Franz Liszt. RESTR – Restrictions expressed as constraints (element in set).

spatial-in-on-rel represents a spatial relationship that can be either *in* or *on*. This relation contains only some selected semantic roles arranged together in a flat format for the sake of clarity. In *spatial-in-on-rel* the values of the attributes P-SUBJ and P-COMP – linked to "neighbour" and "floor", respectively – make sure that an interpretation like "the floor in the neighbour" is not allowed. The HPSG parser will make the values of proto-role attributes be

identical to the semantic content of the actual subject/complements. For determining the actual preposition, the CONCEPT(UALIZATION) in each language is important. This feature is supposed to be encoded within an interlingual semantic hierarchy which contains language-specific conceptualizations as well: In this example, it is assumed that "neighbour" is a *3d-entity* in all considered languages. The CONCEPT of "floor' accounts for the selection of different prepositions: For German, only the typed feature structure (TFS) for *in-spatial*, given in (10), matches with all CONCEPT specifications – *3d-spatial-receptacle* in this case. For English, the same holds for *on-spatial* which unifies with *2d-spatial-plane.*

(39)

$$\begin{bmatrix}
\text{HANDEL} & \textit{handle} \\
\text{INDEX} & \boxed{5} \\
\text{LISZT} & \left\langle \begin{bmatrix} \textit{neighbor} \\ \text{HANDEL} \quad \boxed{1} \\ \text{INST} \quad \boxed{4} \\ \text{CONCEPT}|\text{LANG} \quad \textit{3d-spatial-entity} \end{bmatrix}, \begin{bmatrix} \textit{spatial-in-on-rel} \\ \text{HANDEL} \quad \boxed{2} \\ \text{EVENT} \quad \textit{state}\boxed{5} \\ \text{P-SUBJ} \quad \boxed{1} \\ \text{P-COMP} \quad \boxed{3} \end{bmatrix}, \right. \\
& \left. \begin{bmatrix} \textit{floor} \\ \text{HANDEL} \quad \boxed{3} \\ \text{INST} \quad \boxed{6} \\ \text{CONCEPT} \quad \begin{bmatrix} \text{ENGLISH} & \textit{2d-spatial-plane,} \\ \text{GERMAN} & \textit{3d-spatial-receptacle} \end{bmatrix} \end{bmatrix} \right\rangle
\end{bmatrix}$$

Ambiguous compounds can be represented using MRS – making this approach more suitable than standard HPSG parsing & structure sharing using only non-ambiguous head (corresponding to P-SUBJ) and modifier roles (corresponding to P-COMP) without any relations. There are several further possible sources of ambiguity: Incomplete lexical specifications, ambiguous words or several possibilities of lemmatization: "Kulturinfiltration, Bohrersatz, Urinstinkt".

A possible MRS structure for "Bohrersatz" ("Bohr-ersatz": "Ersatz fürs Bohren"; drilling replacement and "Bohrer-satz": "Satz von-Einheit Bohrer"; set of drills) is:

(40)

$$\begin{bmatrix}
\text{ORTH} & \text{"Bohrersatz"} \\
\text{HANDEL} & \text{handle} \\
\text{INDEX} & \boxed{9} \\
\text{H-CONS} & \{\boxed{9} \in \boxed{2} \cup \boxed{12}\} \\
\text{LISZT} & \left\langle \left\langle \begin{bmatrix} \textit{bohren} \\ \text{HANDEL} \quad \boxed{1} \\ \text{INST} \quad \boxed{4} \end{bmatrix}, \begin{bmatrix} \textit{fuer-rel} \\ \text{HANDEL} \quad \boxed{2} \\ \text{P-SUBJ} \quad \boxed{3} \\ \text{P-COMP} \quad \boxed{1} \end{bmatrix}, \begin{bmatrix} \textit{Ersatz} \\ \text{HANDEL} \quad \boxed{3} \\ \text{INST} \quad \boxed{5} \end{bmatrix} \right\rangle, \right. \\
& \left. \left\langle \begin{bmatrix} \textit{Bohrer} \\ \text{HANDEL} \quad \boxed{11} \\ \text{INST} \quad \boxed{14} \end{bmatrix}, \begin{bmatrix} \textit{von-Einheit-rel} \\ \text{HANDEL} \quad \boxed{12} \\ \text{P-SUBJ} \quad \boxed{13} \\ \text{P-COMP} \quad \boxed{11} \end{bmatrix}, \begin{bmatrix} \textit{Satz} \\ \text{HANDEL} \quad \boxed{13} \\ \text{INST} \quad \boxed{15} \end{bmatrix} \right\rangle \right\rangle
\end{bmatrix}$$

Some compounds are only ambiguous in the preposition that their constituents are combined by: “Oberflächenvertiefung” (an/von: a deepening at the surface; deepening of (entire) surfaces) “Luftkühlung” (von/durch: cooling *of* air; cooling *by* air) “Raketenantrieb” (von/durch: propulsion of rockets; propulsion by rockets). When the following words are being o’§ified as part of a compound, they mostly introduce ambiguity of this kind: “-betrieb” (working/operation); “-steuerung” (steering; control); “-druck” (pressure); “-leiter” (manager; conductor); “-lösung” (solution); “-basis” (basis); “-schwimmer” (swimmer; floater); “-schub” (thrust); “-transport” (transport); “-hülle” (hull); “-regelung” (regulation; prepositions: with, by, at/near, for).

Other compounds are made up of ambiguous constituents combined by an ambiguous relation: “Schulbildung” (construction/formation of schools; education by schools); “Knotenregeln” (rules for forming knots; rules for the behaviour at knots /crossings / for calculating the current at electric knots); “Expertenrat” (aus/von: council of experts; advice by experts):

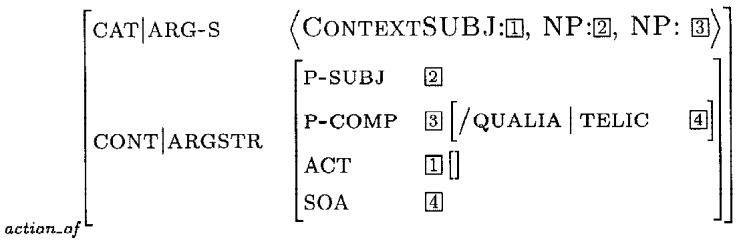
(41)
$$\left[\begin{array}{ll} \text{ORTH} & \text{“Expertenrat”} \\ \text{HANDEL} & \text{handle} \\ \text{INDEX} & \boxed{9} \\ \text{H-CONS} & \left\{\boxed{9} \in \boxed{3} \cup \boxed{4}\right\} \\ \text{LISZT} & \left\langle \left[\begin{array}{ll} \textit{council} & \\ \text{HANDEL} & \boxed{1} \\ \text{INST} & \boxed{6} \end{array}\right], \left[\begin{array}{ll} \textit{advice} & \\ \text{HANDEL} & \boxed{2} \\ \text{INST} & \boxed{7} \end{array}\right], \left[\begin{array}{ll} \textit{of-rel} & \\ \text{HANDEL} & \boxed{3} \\ \text{P-SUBJ} & \boxed{1} \\ \text{P-COMP} & \boxed{5} \end{array}\right], \left[\begin{array}{ll} \textit{by-rel} & \\ \text{HANDEL} & \boxed{4} \\ \text{P-SUBJ} & \boxed{2} \\ \text{P-COMP} & \boxed{5} \end{array}\right], \left[\begin{array}{ll} \textit{experts} & \\ \text{HANDEL} & \boxed{5} \\ \text{INST} & \boxed{8} \end{array}\right] \right\rangle \end{array}\right]$$

Semantic information should be structured in such a way that it is easy to find appropriate relations holding between the constituents of compounds. In the extreme case, this could also mean that constituents do explicitly specify the relations that could hold between them. “Stahlkocher” can mean “Kocher des-Objektes Stahl” (person cooking steel) and “Kocher aus Stahl” (boiler made of steel). “Kocher” as person is a regular sense extension of “Kocher” as tool/machine or “kochen” (to cook) as activity: A person working with the tool/machine or performing the action. In both cases the person is characterized as performing an activity. If the base verb of the action can have an object, then the person performing that action can have the same object as P-COMP value in a compound: “Kocher des-Objektes Stahl” (person cooking steel) in the example. The material of which persons are made is known: Flesh, bones, but normally not steel. “Stahlkocher” can’t be a “person for cooking who is made of steel”: The unification of the MATERIAL values “steel” and “flesh” fails which would be required for the relation m*ade-of*. However, this unification would succeed for “Kocher” as device (“boiler”). Since “Holz” (wood) is a material that under normal circumstances doesn’t melt and isn’t eaten, “Holzkocher” can only be a “Kocher-Gerät betrieben-mit/bestehend-aus Holz” (boiler driven by/made of wood). Here “Kocher” might also be a puppet or statue but not a person. From the fact that wood catches fire when it is too hot but produces heat required for cooking, relation *driven-by* can be preferred to *made-of*. If these facts are not known, the alternative remains to introduce priorities in connection with

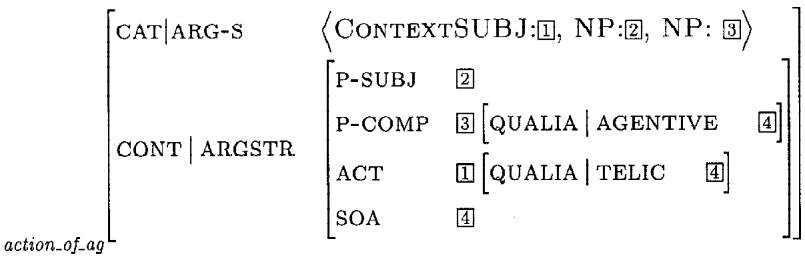
hierarchical relations – using e.g. the CLOS (Common Lisp Object System) class precedence algorithm. In the example, relation *driven-by* could generally be preferred to *made-of* – e.g. by making the first a subtype of the latter or by introducing a subtype of both which specifies its supertypes in the desired order.

An interesting case is "Buchbeginn" which can mean (a) physical beginning of a book, (b) beginning the action of reading a book or (c) beginning the action of writing a book. Distinguishing (a) from (b/c) can be done with the lexical information that "Beginn" is polysemous: It can denote a physical object (spatial beginning) or the starting point of an action (temporal beginning; relation *action_of*). Separating (b) from (c) can be done with the help of some ideas of Pustejovsky applied to the specification of *action_of:*

(42)



(43)



(44) Der Autor (ACT) schreitet zum (P-COMP:) Buchbeginn (P-SUBJ) – dem Schreiben (SOA) (The author (ACT) begins (P-SUBJ) the book (P-COMP): writing (SOA)).

(45) Der Leser (ACT) schreitet zum (P-COMP:) Buchbeginn (P-SUBJ) – dem Lesen (SOA) (The reader (ACT) begins (P-SUBJ) the book (P-COMP): reading (SOA)).

The relation (42): *action_of* (complement per default telic) with its subtype (43): *action-of-ag* (complement definitely agentive) simulate *type coercion* by *default unification*: If – as specified in (43) – the AGENTIVE quale[1] of the *action-of-ag* relation's P-COMP value[2] is identical to the ACT(OR)'s TELIC quale[3], then the state of affairs argument (SOA) must be equal to that value which is in common. This overrides the default given in CONT | ARGSTR | P-COMP of (42). The defeasible part of this default[4] specifies that the SOA value should be

1 Things involved in active creation; writing in (44) and (45)

2 book in (44) and (45)

3 The purpose of an object; reading/writing is the reader's/author's telic role

4 Syntactically introduced by "/"

equal to the telic role of the P-COMP's TELIC quale. This assumption remains intact when analyzing sentence (45) while it is overridden for sentence (44) since relation *action-of-ag* (43) is more appropriate than *action-of* (42).

CONTEXTSUBJ represents a contextual subject like reader or author that may or may not be present within the actual sentence.

Johnston and Busa (1996) use Pustejovsky's qualia structure to explain several compound types: "bread knife" is a "knife" for cutting "bread" – cutting is taken from the TELIC role of knife.

"lemon juice" is "juice" created by squeezing "lemons" with squeezing taken from the agentive quale of the head "juice".

"glass door" is a "door" consisting of "glass" – the constitutive quale being the modifier.

Busa & Johnston use lexical rules to state how the qualia structure of the compounds is generated from that of its constituents. A similar effect is achieved here by finding relations – underlined above (for, created-by, consisting-of, etc.). This search succeeds exactly in those cases in which both compound constituents involved satisfy all syntactic and semantic constraints imposed by the relation. One advantage of this approach is that all kinds of relations between the nouns in compounds are supported – not only those corresponding to the four quales. Furthermore it is possible to express ambiguity using MRS and the scheme to find relations is more general and flexible: Not the lexical data induces the relation but rather the constraints expressed in each relation induce the compatibility with the constituents of a compound.

Probabilistic approaches to compound analysis have been presented in Wu (1990). Other approaches are Nirenburg (1993) and ten Hacken (1994).

5. Conclusion

Similar conceptualizations in terms of preposition usage exist in quite different domains and across languages. This suggests that basic spatial concepts can be generalized to describe even abstract terms like feelings. These ideas are easy applied to foreign language teaching to build up a linguistic instinct. A very similar notion of linguistic instinct is also desirable in the field of computational linguistics: machine translation, disambiguation, selectional restrictions and compound semantics benefit largely from such a semantic account. This approach can be used as alternative or as complement for statistical techniques. The semantic approach can still help where statistical methods fail because of the data sparseness problem. On the other hand, statistical methods could be used to select from alternatives that persist after syntactic and semantic analysis.

A subject to *further research* is the selection of adjectives: Positive adjectives are often associated with a positive value on the *y*-axis ("hohe Anerkennung") while negative adjectives realize the opposite direction: "tiefer Haß", "tiefe Eifersucht". For the same reason, these concepts are used in conjunction with "unter" (under) or adjectives like "tief" (deep) or "hoch" (high). Examples for their spatial concepts are "hoher Tisch" (high table) and "tiefes

Loch" (deep hole). Unfortunately, other inference chains cause many exceptions to this rule: E.g. the same adjectives "tief" (deep), "gering" (little), "hoch" (high), etc. are used to characterize if feelings are superficial.

The *main advantages* of this approach are: Drastically less formal (lexical) specification effort by humans is necessary to represent entire word fields. More cross-linguistic commonalities are identified and exploited. Many instances of creative language use can be explained. Better disambiguation and error correction are possible. The lexical semantic analysis of compounds is more precise. Irresolvable ambiguities are maintained without introducing spurious ambiguities by using MRS representation. Ways for replacing Pustejovsky's type coercion by (more efficient & simple) default unification and for introducing an ordering on (prepositional) relations are shown.

We would like to thank Ken Litkowski (CL Research) for helpful comments and Tony Davis (CSLI) & Stefan Müller (DFKI) in addition to that for making some of their publications in press available to us.

References

Burns, K. J. and Davis, A. R. (1999): Building and maintaining a semantically adequate lexicon using Cyc. In: Viegas, E., editor, Breadth and Depth of Semantic Lexicons. Kluwer Academic Press. In press.

Copestake, A. (1992): The representation of lexical semantic information. PhD thesis, University of Sussex. Cognitive Science Research Paper CSRP 280 Available: http://www.cl.cam.ac.uk/Research/NL/acquilex/papers.html [1997, Jan. 04] Contains (among other things) revised material from acq1wp8, acq1wp21/22, acq1wp36, acq1wp37 and a couple of other papers which are in Briscoe et al, Inheritance, Defaults and the Lexicon.

Copestake, A. (1993): The complete LKB, ACQUILEX-II deliverable, 3.1. Technical report no. 316, University of Cambridge Computer Laboratory. Available: http://www.cl.cam.ac.uk/Research/NL/acquilex/papers.html [1997, Jan. 04]. This has all the formal stuff about the ACQUILEX LKB system.

Copestake, A., Flickinger, D., and Sag, I. A. (1997): Minimal recursion semantics: An introduction. Available: ftp://ftp-csli.stanford.edu/linguistics/sag/mrs.ps.gz [1997, Dec. 05].

Davis, A. R. (1996): Linking and the hierarchical lexicon. PhD thesis, Stanford University. Available: http://www-csli.stanford.edu/users/tdavis/ [1998, Jan. 05].

Davis, T. (1999): Types and Constraints in Lexical Semantics and Linking. Hypothetical publisher: CSLI publications, Stanford, USA. In preparation.

Habel, C. (1989): ZWISCHEN-Bericht. In: Habel, C., Herweg, M., and Rehkämper, K., editors, Raumkonzepte in Verstehensprozessen. Interdisziplinäre Beiträge zu Sprache und Raum, pages 37-98.

Herweg, M. (1989): Ansätze zu einer semantischen Beschreibung topologischer Präpositionen. In: Habel, C., Herweg, M., and Rehkämper, K., editors, Raumkonzepte in Verstehensprozessen. Interdisziplinäre Beiträge zu Sprache und Raum, pages 99-127.

Johnston, M. and Busa, F. (1996): Cross-linguistic semantics for complex nominals in the generative lexicon. In: AISB Workshop on Multilinguality in the Lexicon, Sussex, England. Available: ftp://ftp.cs.brandeis.edu/pub/research/johnston/aisbpaper.ps [1997, Jan. 05].

Lenat, D. B. and Guha, R. V. (1990): Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project. Addison-Wesley, Reading, Massachusetts.

Nirenburg, S. (1993): Lessons from PANGLOSS. In: Proceedings of the Sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Panel on Natural Language Processing for Information Management, page 225.

Pustejovsky, J. (1995): The Generative Lexicon. MIT Press.

ten Hacken, P. (1994): Defining Morphology: A Principled Approach to Determining the Boundaries of Compounding, Derivation, and Inflection. Olms, Hildesheim. 376 pp., ISBN 3-487-09891-1, DM 68.

Trujillo, A. (1995): Lexicalist Machine Translation of Spatial Prepositions. PhD thesis, Computer Laboratory, University of Cambridge, UK. Available: http://admin.ccl.umist.ac.uk/staff/iat/docs/publications.html [1998, Jan. 20].

Wu, D. (1990): Probabilistic unification-based integration of syntactic and semantic preferences for nominal compounds. In: 13th International Conference on Computational Linguistics (Coling90), pages 413-418, Helsinki.