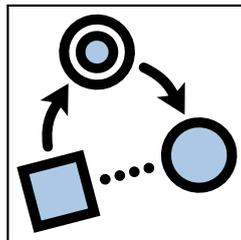


Collaborative Research Center 991

**The Structure of Representations in Language, Cognition, and
Science**

Heinrich-Heine-Universität Düsseldorf



SFB 991

Funding proposal
2/2015 – 1/2019



Proposal for the continuation of Collaborative Research Centre 991

“The Structure of Representations in Language, Cognition, and Science”

funded since

01 July 2011

for

2015/2 – 2016 – 2017 – 2018 – 2019/1

Coordinating university:

Heinrich-Heine-Universität Düsseldorf
Universitätsstr. 1
D-40225 Düsseldorf

Coordinator:

Prof. Dr. Laura Kallmeyer
Institut für Sprache und Information
Heinrich-Heine-Universität Düsseldorf
Universitätsstr. 1
D-40225 Düsseldorf
Phone: +49 (0)211 81 13899
Fax: +49 (0)211 81 11325
Email: kallmeyer@phil.uni-duesseldorf.de

Office:

Lena Hierl
Sonderforschungsbereich 991
Heinrich-Heine-Universität Düsseldorf
Kruppstraße 108 / Geb. 46.21
40227 Düsseldorf
Phone: +49 (0)211 81 12959
Fax: +49 (0)211 81 03170
Email: sfb991@phil.uni-duesseldorf.de

Düsseldorf,

Prof. Dr. Laura Kallmeyer (Coordinator CRC 991)

Düsseldorf,

Prof. Dr. Anja Steinbeck (President HHU)

3.1 About Project C10

3.1.1 Title: A frame-based analysis of adjective noun combinations

3.1.2 Research areas:

104-01 Allgemeine und Angewandte Sprachwissenschaften, semantics, computational linguistics

3.1.3 Principal investigator(s)

Löbner, Sebastian, Prof. Dr., 04.04.1949, German
 Inst. f. Sprache und Information
 Heinrich-Heine-Universität Düsseldorf
 Kruppstr. 108
 40227 Düsseldorf
 Phone: 0211 8113399
 Email: loebner@phil.hhu.de

Petersen, Wiebke, Jun.-Prof. Dr., 05.07.1972, German
 Inst. f. Sprache und Information
 Heinrich-Heine-Universität
 Universitätsstr. 1
 40225 Düsseldorf
 Phone: 0211 8115295
 Email: petersen@phil.hhu.de

Do the above mentioned persons hold fixed-term positions? Yes

Sebastian Löbner: End date 30.06.2015. Further employment is planned until the end of the second funding period.

Wiebke Petersen: End date 02.11.2016. Further employment is planned until the end of the second funding period.

3.1.4 Legal issues

This project includes

1.	research on human subjects or human material	no
2.	clinical studies	no
3.	experiments involving vertebrates	no
4.	experiments involving recombining DNA	no
5.	research involving human embryonic stem cells	no
6.	research concerning the Convention on Biological Diversity	no

3.2 Summary

The goal of the project is a semantic analysis of the compositional interaction of attributive adjectives with nouns in German and English. Rather than aiming at a classification of adjectives, the focus is on conceptual mechanisms of combining the frames of A and N into one. The application of frame theory opens new prospects for long-known semantic problems. In particular it undertakes to analyze adjectival modification on the basis of decomposing the lexical meanings of both the adjective and the noun by representing the relevant aspects of lexical meanings in Barsalou frames. The project will transcend the present state of the art in several regards. (1) Basing the analysis of the lexical decomposition on noun and adjective meanings, it will be able to provide a more explicit picture of the mechanisms of composition, along with analyzes for cases which hitherto lack a plausible treatment. (2) Linking the research to previous frame analyzes of the semantic mechanisms of nominal compounding will contribute to a general picture of modification that also encompasses nominal compounds. (3) Merging the analysis of adjective noun combinations with the theory of nominal concept types will lead to new questions and insights. The data to be analyzed will be provided by comprehensive statistical corpus investigations for German and English, starting with the determination of statistical clouds of adjectives and nouns in combination. In total, the approach will hopefully be able to arrive at a picture of the sub-compositional nature of adjectival modification which is at the same time more differentiated, more explicit, and more comprehensive than other current approaches.

3.3 Research Rationale

3.3.1 Current state of understanding and preliminary work

The phenomena

Adjectives and nouns can be combined in three ways:

- (1) A attribute of N within the NP: [the [black [pen]]]
- (2) lexicalized A+N phrase *black eye, black market, black widow*
- (3) A N compound of type N *blacksmith, blackboard, blackbird, blackmail*

The project is confined to the study of type 1, focusing on the compositional mechanisms involved. A+N phrases present a challenging complex of grammatical and semantic phenomena. Grammatically, adjectives obviously do not form a uniform part of speech. While a noun can be combined with an open number of adjectives, their relative order is not free, and different subtypes of adjectives relate to different functional layers within the NP.

If we consider A+N phrases that consist of just one A and one N, there is a striking manifold of semantic mechanisms, because all subtypes of A can be directly combined with N. If we were to also include productive types of lexicalized A+N phrases and AN compounds, even more mechanisms would enter the scene; see Schlücker (2014) for an investigation of German. In all cases, however, the semantic contribution of A can be considered a modification of the meaning of N, in a sense yet to be discussed below.

Relational A+N. Starting with those adjectives that are most closely combined with N, there is a first layer of A, commonly referred to as ‘relational’, as instantiated in the examples in (4).

- (4) a. nuclear power plant, parental care, musical instrument
- b. Ukrainian crisis, fatherly advice, annual report

Relational A+N use a (mostly) denominal A and express a relation of N to what the root of A denotes. In most cases, relational adjectives are used for ‘classificatory’ modification; some authors just speak of ‘**classificatory**’ adjectives (Morzycki, to appear). But it has been pointed out that there are two applications of relational A+N, one classificatory, one not (Schlücker, 2014). In the non-classificatory use, exemplified in (4b), the adjective relates the actual referent of the noun to something denoted by the root of A. In cases like *Ukrainian crisis*, this may lead to an expression that, in a particular context, denotes a unique referent. In the classificatory function – illustrated in (4a) – the adjectives are regarded to serve the formation of a term that denotes a class of cases derived from the general class denoted by N alone. Relational adjectives are mostly denominal, but there are also derivations from other parts of speech such as German deadverbial *morgig, heurig, baldig, hiesig*, or Russian deverbal *пишущая* (from *писать* ‘write’) in *пишущая машинка* ‘typewriter’, lit. ‘writ-ish machine’. Denominal relational adjectives modify the meaning of N by relating the meaning of their nominal root to their modifiee. In principle, there is the possibility of using NN compounds instead of denominal A + N, cf. German *Atomkraftwerk, Elternsorge, or Musikinstrument*. In English and in German, both options exist, sometimes with the same A or the same N. The semantic analysis of the compounding mechanisms of NN compounds can therefore be expected to shed light on the compositional mechanisms involved in A+N phrases with a denominal relational A. (See ‘Preliminary work by the PIs’)

Event-related A+N. There is a further class of A+N where the interaction of A and N requires deeper decomposition: The type represented by *safe driver* or *good chef*. In these cases, the noun is either deverbal (*driver*) or else semantically related to events (like *chef* is to cooking) and the adjective has the same effect as an adverb would have on a verb denoting the event. This type of A may be separated from N by intervening classificatory or non-classificatory A; for example ‘good French chef’ is a possible string and has both readings for *French*, classificatory and non-classificatory. What unites this class of cases with the former is the close and varied interaction of the meanings of A and N where one of them derives from a third source (deverbal N or denominal A). We will refer to adjectives in this function as ‘event-related’. The nouns they are combined with need not be deverbal – we have both *lousy teacher* and *lousy professor, good cook* and *good chef, fast driver* and *fast car*.

Property A+N. The literature on the order of adjectives (see Svenonius, 2008, for a survey) agrees in the observation that the layer(s) first following classificatory adjectives are non-gradable adjectives relating to material (*wooden*), origin (*Italian*), color, and shape, in this order, starting from N. These adjectives clearly specify one attribute of the N referent (MATERIAL, ORIGIN, COLOR, SHAPE). It is only after these adjectives that the prototypical (and best studied) adjectives enter the stage: gradable property adjectives such as dimensional adjectives *big, long, old, hot, and expensive*, and non-dimensional adjectives like *good,*

healthy, and *clean* (for the distinction see Bierwisch 1989 who calls non-dimensional adjectives ‘evaluative’). Following Morzycki (to appear), we will refer to all these types of adjective as ‘property adjectives’. Unlike classificatory adjectives, property adjectives can also be used predicatively. The respective status of event-related adjectives is a matter of debate. Some consider material-adjectives such as *wooden* as relational adjectives (see Trost, 2006, p. 133 for references). This would explain the fact that these A lack predicative use.

There are further types of A+N which will not be discussed since the research in the project will be confined to the three types mentioned so far. The further types include:

- Modal A+N. Adjectives such as *former*, *future*, *would-be*, *potential*, *alleged*, *fake* impose some kind of temporal or modal quality on the predication of N. They often lead to ‘privativity’, i.e. to reference of A+N to cases that are not, or may not be, ‘N’.
- Quantitative A+N. Numerals, both cardinal and ordinal, along with vague quantity specifications such as *many*, *much*, *few*, *more*, *enough*.
- Ordering A+N. Adjectives that place the referent of the noun in some order. The class includes ordinals, superlatives, and adjectives such as *last*, *next*, *new*, *old* or *favorite*.
- Further layers of operators within the NP. These include adjectives for discourse location (*latter*, *above-mentioned*), as well as demonstratives, articles, and quantifiers proper (*every*). Some of these are adjectival such as German *[die] ganzen*. Others are traditionally not considered A.

Current state of understanding

Logical analysis. The current state of the art in formal semantics with respect to adjectival and adverbial modification is discussed in Morzycki (to appear) and McNally (to appear). Both sources treat adjectival and adverbial modification, but exclude modification by compounding. The sources essentially agree in defining modification as a mechanism by which a modifier is applied to, or combined with, a modifiee, resulting in a complex phrase of the same logical type as the modifiee. The current approaches to the phenomenon of adjectival modification are in terms of semantic subclasses of adjectives with different logical properties, among them:

Intersective adjectives, such as color adjectives denote a property and hence are of logical type $\langle e, t \rangle$ as are their operand nouns. Combining A and N has the effect of forming an expression whose denotation is the set-theoretical intersection of the sets denoted by A and N. Equivalently, the composition of N and A, both type $\langle e, t \rangle$, can be described as truth-conditional conjunction. Intersective adjectives comprise non-gradable and gradable property adjectives. They can be used attributively and predicatively.

Subjective adjectives effect a restriction on the denotation of the noun, but they are not intersective. Alternatively, intersective adjectives can be considered a special subclass of subjective adjectives. Of the classes mentioned above, subjective adjectives include classificatory adjectives and event-related adjectives. Subjective adjectives can be used predicatively, but the predicative construction is not necessarily equivalent to the attributive construction: a ‘beautiful dancer’ is not necessarily beautiful, etc. Dimensional adjectives such as *big*, *long*, *expensive* are controversially discussed. As they involve a comparison class which may depend on the noun, they are considered subjective by some, while others argue that they can be considered intersective (cf. the discussion in Morzycki, to appear, s. 2.2.3).

Intensional adjectives include the modal adjectives mentioned above. They invest the modifiee with a temporal, modal, or epistemic qualification.

Open problems. Concerning the work in formal semantics, it is to be noted that the distinction of logical types of nouns (cf. Löbner, 2011), such as relational nouns or inherently unique nouns, has not been sufficiently taken into account. Relational adjectives may shift relational to absolute nouns by argument saturation. Conversely, an adjective like *favorite* induces or requires an additional argument of the noun. Order adjectives shift non-unique CNPs to unique CNPs. Property adjectives, if restrictive, require non-unique nouns. Inherently unique nouns, on the other hand, combine naturally with intensional adjectives such as *former*, *elected* etc. Relational nouns may be open to adjectival attributes adding to the relational concept (*left eye*, *younger sister*). – Another open problem is to account for, and model, the obvious semantic asymmetry between modifier and modifiee. Intuitively, the modifier adds a limited contribution to the meaning of the modifiee, its weight being much less than the weight of the modifiee. For example, in the prototypical case of property adjectives, the modifier adds just one property to the meaning of the modifiee which may contain many more conditions of its own. The intersective analysis treats both as of equal weight. Approaches in terms of functional application likewise are unable to explain the difference. This shortcoming is inevitable for an approach that ignores lexical meaning for all but logical type; logical type assignment doesn’t determine the degree of conceptual complexity.

Relational adjectives and event-related adjectives. Relational adjectives are hard to handle in formal semantics because there is no principled way of getting at the decompositional level where the meanings

of A and N would interact. McNally & Boleda (2004) proposed to analyze certain cases of classificatory relational A+N as A applying to N at the kind level of denotation rather than at the level of single referents. For example, they describe the meaning of *x is a technical architect* as (informally) 'x realizes a kind k that is an architect and is technical'. This analysis involves an intersective use of the adjective at the kind level. Another strategy has been employed in computational linguistics by using semantic primitives for the relations involved, e.g. a primitive instrument for characterizing relational A+N where N is linked as an instrument to something specified by A (Ó Séaghdha & Copestake, 2009). Event-related A+N such as *beautiful dancer* have been approached in a similar way in Davidsonian event semantics, for example analysing *beautiful dancer* as $\lambda x \exists e[\text{dance}(e) \wedge \text{AGENT}(e)=x \wedge \text{beautiful}(e)]$ (Larson, 1998).

Open problems. The approach taken by Larson to event-related A+N appears viable and will be further examined in the project. However, the two mentioned treatments of relational nouns are both problematic. The kind-level approach fails to generalize to other relational A+N cases. For example, a *linguistic theory* is not a theory of the linguistic kind, *German linguistics*, at least in one reading, is not linguistics of the German kind, and so on. Relating the adjective as a predicator to a kind variable only shifts the problem of coming up with the exact relation required. A different kind of criticism applies to the semantic primitives approaches: they add a conceptual component to the meaning of A+N, where in fact this component is often retrievable from the lexical meaning of N.

Order of adjectives. Syntactic accounts such as Cinque (2010) assume that there is a universal hierarchy of functional projections of the noun which provides a structural specifier position for each type of adjectival modifier. Purely semantic accounts aim at explaining the order restrictions in terms of the meanings of subclasses of adjectives, for example by scope relations or by independent principles (such as: adjectives denoting more permanent properties are placed closer to the noun, Stassen, 2003). Mixed approaches assume a smaller hierarchy of structural positions; in a given position there may be more than one adjective, e.g., two gradable property adjectives, whose relative position would be determined by their semantics.

Open problems. There is strong semantic evidence for the distinction of functional layers within the NP (Löbner, to appear in 2015; Svenonius, 2008) but we do not think that these layers are as numerous and as fine as is postulated by the extreme syntactic approach taken by Cinque. Rather we would assume that the syntactic combination of A+N at a given functional layer is a sub-compositional construction, i.e. a syntactically uniform construction that hosts a variety of semantic modes of composition (Löbner, 2012a). This would hold in particular for the innermost layer of relational adjectives. We formerly established that the gradation of verbs (in fact a modification of closely related type) is a case of sub-compositionality, and A+N is a very plausible candidate for another, or rather, a set of more such constructions.

Current state of understanding: Decomposition. It may be assumed that the meaning of adjectives is conceptually leaner than that of verbs and, even more than that of nouns. The meaning of prototypical adjectives such as *red*, *big* or *good* appears to capture just one aspect of the adjective argument. This is particularly plausible for dimensional adjectives. This idea has been spelled out in various ways and there are fully decompositional analyzes of dimensional adjectives, among others by Bierwisch (1987); Löbner (1990), and Kennedy (1999). These accounts are restricted to gradable adjectives, excluding non-gradable classes such as color, size, shape, or material adjectives. The analyzes also do not apply to non-dimensional gradable property adjectives such as *clean*, *good*, *healthy*. It was recently suggested by Sassoon (2013) that these adjectives involve more than one dimension. Concerning the noun, there are decompositional analyzes for event-related A+N with deverbal noun (Larson, 1998). As a further step towards N decomposition, Morzycki (2005) proposes that nouns might be associated with bundles of features that correspond to the dimensions specified by certain adjectives. And, of course, there is the qualia approach by Pustejovsky (1995), which also has been applied to A+N modification such as in *fast car*.

Open problems. The dominating treatments of dimensional A (Bierwisch 1987 and similarly Kennedy 1999) analyze positives underlyingly as comparatives with a comparison degree to be retrieved from context, e.g. as a statistical standard for the respective noun class, or to be explicitly specified with a degree adverb. These treatments are contrary to traditional wisdom, which considers the positive of A as a one-place predicate term (except for two-place adjectives such as *keen (on)*) and they contradict the fact that in the vast majority of languages comparative forms, if they exist, are morphologically more complex than positive forms. The decompositional analysis of event-related adjectives proposed by Larson has problems, too: it cannot be easily applied to A+N phrases with nouns that are not morphologically related to verbs. A problem of the feature bundle approach proposed by Morzycki (2005) is the assumption that each feature comes with a variable to be bound, rendering ordinary nouns as multi-place predicate terms without a corresponding syntax.

Corpus data. There is a corpus-based study of German adjectives by Trost (2006). Trost presents a comprehensive classification of 213 high-frequency German adjectives with a total of 1088 polysemous readings. He arrives at a detailed classification based on semantic, morphological (gradability) and gram-

matical (attributive vs. predicative use) criteria. The semantic distinctions are fine-grained, but they are not elaborated into explicit formal analyzes. There is also emerging research on A+N phrases in the field of Distributional Semantics. Vecchi, Zamparelli & Baroni (2013) present a study on the relative order of prenominal adjectives in English A+A+N phrases. They show that there are two classes of A; one with fixed relative order, and the other with free relative order. They define a measure for the strength of the modification by an adjective by measuring the angle between the vectors for N and A+N. This measure might be a useful indicator for distinguishing types of adjectival modification. Van de Cruys (2006) reports results of a semantic clustering study using vector space measures for Dutch adjectives and nouns. Clusters are built on the basis of the nouns co-occurring with an adjective and vice versa. The resulting noun clusters are evaluated with respect to EuroWordNet relations like 'synonyms' and 'hypernyms'. Van de Cruys reports a maximal precision of 42.5% and a recall of 8% when comparing his clusters with those in EuroWordNet, while intra-cluster similarities of between 50 and 60% are reached when the semantic similarity measure of Wu and Palmer is applied to the noun clusters.

Open problems. Although Trost (2006) is a rather detailed corpus-based study, it is far from covering all relevant data. One shortcoming of the study is its restriction to highly frequent adjectives; this is a principal problem with distributional approaches, too. Rare A+N combinations, even hapax legomena, are valuable data for the type of phenomena we are after because they are very likely to instantiate *productive* types of modification (while frequent combinations are prone to being lexicalized or at least conventionalized). Since the project aims at finding the productive patterns of A+N modification, in particular of relational A+N, low frequency data are highly important. Nevertheless, the approaches mentioned provide useful insights. For the automatic clustering approaches the main difficulty is to identify good cluster sizes and to develop appropriate evaluation measures. As the precision and recall values reported in Van de Cruys (2006) demonstrate, using standard partitional and hierarchical clustering methods may not produce optimal clusters. Therefore, the project will also examine and adapt density-based clustering algorithms and compare their performance with hierarchical and partitional methods.

Preliminary work by the PIs

On frames and compounds. The relevant preliminary work comprises research and theorizing on frames as a format of conceptual representation, in particular of lexical meanings (Löbner, 2012b, 2014; Petersen, 2007; Petersen et al., 2008). This work opens up the perspective of modeling the lexical meanings of A and N and their interaction in A+N phrases by means of attributes in frames. In Project C05 in the first funding period, Löbner and his collaborator Daniel Schulzek developed frame analyzes for various types of German NN compounds. First results are published in Löbner (2013, s. 12.3) and in Schulzek (2014). The work proposes analyzes of four compound types, all paralleled by relational A+N phrases:

- | | | | |
|-----|----|--|--------------------------------|
| (5) | a. | argument compound (the modifier specifies an argument of the head) | |
| | | NN <i>air pressure</i> | A+N <i>Russian pressure</i> |
| | b. | value compound (the modifier specifies the value of a frame attribute of the head) | |
| | | NN <i>wood frame</i> | A+N <i>wooden frame</i> |
| | c. | synthetic compound (the modifier specifies an argument of the root verb of the head N) | |
| | | NN <i>taxi driver</i> | A+N <i>nuclear reaction</i> |
| | d. | frame compound (modifier and head both are values of attributes in a bridging frame) | |
| | | NN <i>coffee mug</i> | A+N <i>surgical instrument</i> |

The parallels suggest that it is promising to extend the frame approach to other types of relational A+N phrases as frames are capable of representing an open variety of relations, e.g. all relations corresponding to the attributes concerned, in value and argument compounds and A+N phrases.

For simple property adjectives, the frame approach offers a decompositional analysis: a one-place property adjective can be modelled as a frame with an open argument with one specific attribute and a specification of its value (Petersen et al., 2008). For example, in matrix notation, the meanings of *big* and *woolen* would be as in (6) where x represents the open argument, 'marked ↑' stands for a position on the scale that is marked, in this case, against smaller entities in the domain of the SIZE attribute.

- | | | | | |
|-----|----|-----------------------------------|----|--------------------------------------|
| (6) | a. | <i>big</i> (x): [x.SIZE:marked ↑] | b. | <i>woolen</i> (x): [x.MATERIAL:WOOL] |
|-----|----|-----------------------------------|----|--------------------------------------|

When A modifies N, the x node is unified with the referential node of the N frame, where unification will fail if the referent is of a type for which the attribute is not defined (see Petersen 2007 on the role of types in Barsalou frames). The N frame may come with constraints on the values and the co-variation of values of the attributes of the referent. For example, a 'mouse' frame will provide a constraint on possible, and average, values of the SIZE attribute for mice, depending on their age. This constraint plays a crucial role in calibrating the size scale as to which mice are markedly big or small (where the crucial principle to be

applied is Kamp & Partee's (1995) 'non-vacuity principle' that requires that there be positive and negative, i.e. marked and unmarked, cases possible in the given context). The existence of constraints like these is a further advantage of the decompositional frame approach as it enables one to do without the assumption of some standard provided by the context. These considerations on the frame representation of property adjectives have not yet been worked out, but they arise naturally from cognitive frame theory.

On internal NP structure. In Löbner (to appear in 2015), a model is presented of functional layers of the NP/DP, of the operators and operations at each level, and of the changes to the nominal concept that these operations effect, in terms of the conceptual features relationality, countability, and uniqueness. The model distinguishes the nucleus and seven sequential layers of operation: relation < quality < unit < quantity < ordering < determination < quantification. It allows the allocation of types of adjectival modification at different functional layers and the description of their functional effects. The model is an elaboration of the theory of concept types and determination presented in Löbner (2011) and further applied in projects C01 and C02. Frame theory has so far been developed for the levels *Nucleus* (Petersen, 2007; Petersen & Osswald, 2012) and *Relation* (Petersen & Osswald, 2014). This project proceeds to the level Quality for property and event-related A+N, while the analysis of relational A+N will elaborate the levels Nucleus and Relation. Löbner (to appear in 2015) contains first observations on the interaction of different types of adjectives with nominal concept types and the type shifts effected by adjectival modification, an issue rarely addressed in formal semantics.

On dimensional adjectives and phase quantification. In Löbner (1990, ch. 8) a theory of dimensional adjectives was developed that analyzes these adjectives as instances of phase quantification. This approach is able to avoid the problematic assumption of a degree argument (Bierwisch, Kennedy) not indicated by syntax or morphology. At the same time, it is able to provide the necessary arguments and parameters of analysis if demanded by syntactic additions such as degree phrases or comparative morphology. Naumann (2014) provides a recent proposal for modeling phase quantification within frame theory.

On automatic data collection and classification. For a pilot study, we created a vector space model of German adjectives, whose similarities are measured by the collocated nouns. Data for this model were extracted from a corpus of 100.000 newspaper sentences, which were POS-tagged using the MATE parser (Bohnet & Nivre, 2012). We used the Jaccard measure for calculating the similarities between pairs of adjectives, filtered out high scoring pairs of adjectives, and transformed them into a graph structure. Applying a density-based clustering algorithm to these data reliably detects clusters of adjectives derived from proper names such as a 37-member cluster of city adjectives (*Konstanzer, Stralsunder, Ingolstädter, ...*), but also less obvious groups such as the car adjectives *vorausfahrend, geparkt, gepanzert, etc.*

In the context of a student research project under the guidance of Wiebke Petersen, an additional pilot study on automatic clustering of German nouns and adjectives is being carried out. This study has four main aims: (1) to experiment with different techniques for the automatic detection of adjective-noun pairs, (2) to explore how well algorithms that belong to different families of clustering algorithms perform in our task, (3) to estimate good cluster sizes, and (4) to develop first ideas on how the quality of the automatically gained clusters can be evaluated. (ad 1) Besides using full parsers we experiment with simple pattern recognition techniques on the Google Books Ngram Corpus in order to build up a collection of attested A+N pairings. In a first preliminary step, we collected 361.945 A+N types distributed over 643.309.769 tokens. (ad 2,3) For the clustering step, different standard algorithms have been used (hierarchical clustering, k-means, biclustering), of which, up to now, biclustering with relatively small clusters has provided the most promising results. (ad 4) The results will be evaluated with respect to GermaNet noun and adjective classes.

3.3.2 Project-related publications

- Löbner, S. (1990). *Wahr neben Falsch. Duale Operatoren als die Quantoren natürlicher Sprache*. Tübingen: Niemeyer.
- Löbner, S. (2011). Concept types and determination. *Journal of Semantics* 28:279-333.
- Löbner, S. (2012). Sub-compositionality. In W. Hinzen, E. Machery, M. Werning (eds.), *The Oxford Handbook of Compositionality*. Oxford: Oxford University Press. 220-241.
- Löbner, S. (2013). *Understanding Semantics*. 2nd edition. London, New York: Routledge.
- Löbner, S. (2014). Evidence for frames from natural language. In T. Gamerschlag, D. Gerland, W. Petersen, R. Osswald (eds.), *Frames and concept types. Studies in Linguistics and Philosophy. Vol. 94*. Heidelberg, New York: Springer. 23-68.
- Löbner, S. (to appear 2015). The semantics of nominals. In N. Riemer (ed.), *The Routledge Handbook of Semantics*. London, New York: Routledge.

- Petersen, W. (2007). Representation of concepts as frames. In J. Skilters, F. Toccafondi and G. Stemberger (eds.), *Complex Cognition and Qualitative Science*. University of Latvia. 151-170.
- Petersen, W., J. Fleischhauer, H. Beseoglu, P. Bücken (2008). A frame-based analysis of synaesthetic metaphors. *The Baltic International Yearbook of Cognition, Logic, and Communication*, New Prairie Press 3. The Baltic International Yearbook of Cognition
- Petersen, W. and T. Osswald (2012). A formal interpretation of concept types and type shifts. In K. Kosecki and J. Badio (eds.), *Cognitive Processes in Language*. Frankfurt: Peter Lang. 183-191
- Petersen, W., T. Osswald (2014). Concept composition in frames: Focusing on genitive constructions. In T. Gamerschlag, D. Gerland, R. Osswald, and W. Petersen (Eds.), *Frames and Concept Types. Applications in Language, Cognition, and Philosophy*. Heidelberg, etc.: Springer. 243-266.

3.4 Research Plan

3.4.1 Research questions, aims and hypotheses

The research question of the project is this:

What are the semantic mechanisms of non-lexicalized relational, event-related, and property A+N phrases in English and in German and how can they be modelled in a frame approach to the lexical meaning of A and N and to the compositional mechanisms?

The research plan is based on four working hypotheses:

- Working hypothesis 1:** There is a principal asymmetry between the semantic contribution of modifier and modifiee: the modifier furnishes a minor contribution, by adding, or specifying, one attribute in the modifiee frame. The content of the modifiee concept is preserved in all cases.
- Working hypothesis 2:** The key to an understanding of (i) adjectival modification and (ii) the lexical meaning of adjectives is not a classification of adjectives with respect to their logical properties, but an understanding of the variety of semantic mechanisms in which the meaning of the adjective interacts with the meaning of the noun.
- Working hypothesis 3:** The frame approach is adequate for representing the semantic mechanisms of adjective-noun interaction.
- Working hypothesis 4:** The compositional effect of adjectival modification of any sort is the specification of one attribute in the nominal frame.

3.4.2 Work packages

WP 1 Data collection

WP 1.1 Raw data. In order to arrive at a comprehensive picture of the phenomena, we will extract A+N pairings from large POS-tagged corpora of English and German belonging to different text types. These A+N pairings will be filtered to eliminate pairs with lexicalized meanings. The elimination can be partly based on comprehensive existing lists, and partly performed by statistical methods. A first selection of data will yield the set of those adjectives A which only occur immediately adjacent to an N. This collection will contain all relational adjectives. Among the remaining adjectives, the property adjectives can be recognized automatically by their grammatical behavior, because they are capable of both predicative and attributive use. Within this group, we will distinguish gradable from non-gradable adjectives using automatic methods. There will remain the problem of polysemy. Adjectives are heavily polysemous, and the polysemes often occur in different modifier functions. We are planning to address this problem with methods of distributional semantic analysis by examining the noun clusters with which each of the polysemous adjectives occurs.

WP 1.2 Data postprocessing. (1) POS tagging of existing corpora such as the BNC is often unreliable. For an accurate part of speech recognition postprocessing is needed, such as checking if an item tagged as A is also tagged as N, V, or proper name and cleaning the results accordingly. (2) We will further subject the lists of nouns and adjectives to automatized morphological parsing, in order to identify the roots of derived adjectives (note that probably all relational adjectives are derived) and to obtain more evidence for adjective subclassification; for example, relational adjectives do not take the same affixes of negation as property adjectives. Morphological parsing of the nouns will detect candidates for event-related adjectival modification.

WP 2 Automatic detection of A and N classes

Methods of distributional semantics and automatic clustering such as density-based clustering, biclustering

and multilevel graph partitioning will be used to determine clouds of A+N phrases for the more frequent A and N combinations. Following state-of-the-art approaches in distributional semantics, we will start with vector space representations of the problem. Special attention will be paid to finding an appropriate similarity measure; our current experiments have shown that even a simple measure such as the Jaccard distance, which works with binary features, can produce meaningful results. Given the drawbacks of partitioning and hierarchical clustering methods (esp. regarding predefined cluster numbers, the tendency to produce equally sized clusters, and the lack of an automatic split criterion for hierarchical approaches), our focus will be on density-based clustering algorithms. We will evaluate established methods such as DBSCAN, but also newer approaches such as Markov clustering, which makes use of the weighted structure of the similarity graph. Another focus of the project is the evaluation of the clusters identified. First, they will be compared with existing structured resources such as GermaNet. Second, we will also apply measures of intra-cluster similarity. As we don't aim at reproducing the structures of GermaNet using automatic methods, but rather at finding coherent semantic classes, we expect interesting insights from comparing the optimal clusterings according to the GermaNet similarity and to the intra-cluster similarity.

While these procedures will not capture infrequent combinations, they will nevertheless provide first semantically significant loose groupings. From these clouds we will manually distil semantically more homogeneous classes of A and of N, arriving at a first heuristic subclassification of A, N, and A+N combinations. Results of the semantic analysis in WP 3 will be fed back into WP 2 in order to provide established semantic A and N subclasses as targets to be approximated by automatic clustering.

WP 3 Semantic analysis

WP 3.1 Pre-analytical steps. As a preparation for the analyzes proper, we will isolate hapax legomena A+N phrases, and try to assign the A and the N to the heuristic classes determined so far. As the main aim is not a classification of A or of N, but an understanding of the types of combinations, we will develop a questionnaire for paraphrasing A+N combinations. Developing the questions will be restricted and guided by the gross classification (into relational, event-related and property A+N modifications) and by the heuristic sub-class characteristics.

WP 3.2 Frame analysis. Based on the annotations of the hapax legomena A+N classifications, and on the distinction of the gross A+N classes and the annotated paraphrases, we will turn to the frame analysis of groups of cases, starting from the results obtained in project C05 for nominal derivations and compounds and from natural A classes such as 'ethnic' A like *French*.

WP 3.2.1 Relational A+N. In the present state of our research, we will content ourselves with frame models for representing the overall meaning of A+N. If, for example, A is denominal with root N1, and combined with N2, we will aim at a frame representation that makes explicit how the frame for N1 is integrated in, or with, the frame for N2. We will not attempt to represent the frame A first, the way it derives from N1, and then model its combination with the frame for N2. This appears a reasonable way to proceed given that it is not at all clear if there is a uniform analysis of denominal adjectives which can be upheld in all cases of relational A+N modification in which a particular adjective might engage. Choosing this option means adopting the point of view of construction grammar for this type of modification. However, this does not exclude the possibility of later generalizations which cover more, or perhaps all, existent patterns. The issue whether relational adjectives should be assigned a meaning on their own (rather than a contribution to the construction meaning) raises ontological questions concerning the resulting type of denotations; these ontological questions will be investigated in collaboration with the philosophical project A05.

WP 3.2.2 Event-related A+N. For this type, we will do **joint work with project B09, WP 4.1**. The adjectives relate to an event component in the noun phrase which (at least in many cases) can be modelled as an affordance node in the noun phrase; for deverbal nouns, the affordance can be modelled in terms of an argument attribute of the root verb. Event-related adjectives apply to an event in the same way as manner (and possibly) other types of adverbials. This project will therefore take over and implement analytical results of B09. An empirical question, however, is whether there will always be matching pairs of Adv+V and A+deV N.

WP 3.2.3 Property A+N. Different subclasses of property adjectives require different procedures. For non-gradable property adjectives, we will try to assess whether or not they relate to a unique attribute. This is no easy task, as there is no derivational mechanism as with gradable adjectives that would derive a functional noun for the attribute related (nouns like *color, shape, material, origin* are not deadjectival, unlike dimension terms such as *length*). Thus, assessing the relevant attributes will be an exploration of hitherto little known nominal frame attributes. Gradable dimensional adjectives can be organized in pairs of antonyms, and jointly be assigned to a common attribute/dimension. The expected picture will be much more fuzzy for non-dimensional, 'evaluative' adjectives. This question will be partly addressed by employing annotators.

WP 3.3 Adjectival modification and nominal concept type. The conceptual type of the noun (in terms of inherent uniqueness, relationality, and countability) restricts the range of possible adjectival modification; e.g. property A require non-unique nouns. Some adjectival modifiers shift the conceptual type of the nominal; e.g., superlatives and ordinals yield inherently unique nominals.

WP 4 Theoretical work on frames

WP 4.1 Dimensional adjectives as phase quantifiers in frame theory. The theoretical work will include the implementation and elaboration of the phase quantifier analysis of dimensional adjectives in frame theory. It will further address the question of constraints on, and general aspects of, modeling the combination of A and N frames as well as A and event frames for event-related A+N and of two N frames for denominal relational A+N modification. If possible, we will try to come up with an explanation for the order restrictions observed within the classes of A+N we investigate. A further general question concerns the distinction between classificatory modification and modification at the level of the referent(s) of the nominal. This might include theoretical modeling of reference within frame theory. The topic is directly related to WP 3.1 in project A01 that deals with the different functions, or interpretations, of frames, including the representation of a concept for a category and the description of a particular instance of the category.

WP 4.2 Constraints and prototypes. A further line of theorizing concerns the role of prototypes and probabilistic value constraints in the process of combining the A and N frame. A first point of departure from Barsalou's perspective is the above-mentioned constraints on attribute values and on the interdependence of attributes. In order to predict such constraints for the resulting A+N frame, more elaborate models will be needed like those discussed in project D01. We will closely collaborate with D01 and our designated Mercator fellow Henk Zeevat. From the point of view of linguistic semantics, a basic question for project D01 is for which types of A+N modification the question of attribute value constraints arises, and which attributes are concerned.

WP 5 Towards a theory of modification

WP 5.1 Adjectival modification in general. We will pursue the working hypothesis 4 and thereby provide a model of modification that explains and characterizes the asymmetry of the semantic contributions of modifier and modifiee. If the hypothesis can be corroborated, it will lead to a more general hypothesis on modification as such.

WP 5.2 Comparing adjectival and adverbial modification. In immediate collaboration, the two projects on modification, B09 and C10, will unify their results in order to arrive at central claims for a general theory of phrasal modification. (**WP 5.2 = WP 4.2 in project B09**)

Schedule

2015/2	2016/1	2016/2	2017/1	2017/2	2018/1	2018/2	2019/1
WP 1.1	WP 2						
	WP 2						
	WP 3.1						
	WP 3.2						
WP 4.1			WP 4.2			WP 5	

3.5 Role within the Collaborative Research Centre

The project aims at a substantial contribution to frame theory and its applications. One contribution is at the level of lexical meanings as the project will develop fully decompositional models for adjective meanings (at least for property adjectives) and partially decomposed models for noun meanings. In tandem with project B09, C10 undertakes an extensive study of modification as what could be called 'close composition', i.e. composition at the scale of the interaction of two lexical frames. Unlike composition of phrases such as verb complements with a verb, mechanisms of close composition seem to be more manifold and tuned to finer semantic classes. Thus their analysis offers evidence for the inner structure of the lexical meanings concerned. B09 and C10 investigate the interaction of A and V and A N, respectively. B08 approaches, with different methods, the interaction of V and N. Thus, the three projects scrutinize the pairwise conceptual interaction of the three major lexical categories. They thereby contribute to the general aim of developing a frame account of composition based on decomposed lexical meanings in frame format. Naturally, the development of such a theory has to start out bottom up from the base level of close composition. A second contribution, together with projects B08 and B09 is the use of statistical methods; the project also links the CRC to current trends in linguistics and computation.

Project A01 (Petersen). C10 proceeds in close collaboration with A01 in all principal questions of frame theory and its application. These include in particular the structure of adjectival and noun frames; the theory of frame composition involved in A+N modification; the role and nature of constraints on attributes in frames; the role of type restrictions in frame composition; the distinction of frame modification applying at the general concept level as opposed to the level of description of a single case of reference.

Project A05 (Kann). C10 will discuss with A05 ontological questions regarding the Aristotelian distinction between substances (denoted by nouns) and properties (denoted by adjectives). A promising point of departure is the observation that adjective frames, unlike noun frames, lack a referential node. Immediately related to this observation is the question of the status of the categories defined by adjectival frames in the ontological signature associated with frames in application. The work of Kann (A05) on the medieval theory of the properties of terms is concerned with adjective noun combinations as cases of restriction (restrictio) of a term's contextual reference. This anticipates the notion of 'subsective' adjectives in modern semantics. A05 and C10 can collaborate in representing different modes of such restriction by means of frames, also covering relational adjectives and other variants of 'subsective' adjectives. Certain classes of subsective adjectives, such as the relational adjective in *ontological theory* do not denote properties. This raises the question as to whether these adjectives denote a particular ontological type (a relation?) or should be considered derivative of the underlying noun. The ontological question corresponds to the semantic question as to whether the semantic analysis should aim at a semantic representation of relational adjectives on their own or content itself with a holistic interpretation of the construction 'denominal A + N'.

Project B09 (Löbner & Petersen). C10 proceeds in close collaboration, partly in parallel with the project B09 on verb modification. Parts of the results of B09 can be immediately adopted in C10 for modeling event-related A+N modification. The two projects will also develop common methods of automatic data retrieval and processing. Exploring the two major cases of modification – V modification and N modification – the two projects will jointly work towards a general semantic theory of phrasal modification.

Project C09 (Filip). The project will collaborate with C09 on the issue of countability for adjectives. Adjectives are mass (e.g. color terms) and count (e.g. dimensional A), too; a distinction that is related to the respective positions of the adjectives and to the NP-internal layer they apply to.

Project D01 (Schurz). The project will discuss with D01 and with the Mercator fellow Henk Zeevat the question of the predictability of constraints on attribute values and prototypicality for the results of adjectival noun modification (WP 2 of D01).

Bibliography

- Bierwisch, M. 1987. Semantik der Graduierung. In M. Bierwisch & E. Lang (eds.), *Grammatische und konzeptuelle Aspekte von Dimensionsadjektiven*, 91–286. Berlin: Akademieverlag.
- Bierwisch, M. 1989. The Semantics of Gradation. In M. Bierwisch & E. Lang (eds.), *Dimensional Adjectives*, 71–261. Berlin, New York: Springer.
- Bohnet, B. & J. Nivre. 2012. A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing. In *EMNLP-CoNLL*, 1455–1465.
- Cinque, G. 2010. *The Syntax of Adjectives: A Comparative Study*. Cambridge MA: MIT Press.
- Van de Cruys, T. 2006. Semantic clustering in Dutch. In K. S. et al. (ed.), *Proceedings of the 16th Computational Linguistics in the Netherlands (CLIN)*, 17–32. Amsterdam.
- Kennedy, C. 1999. *Projecting the Adjective. The Syntax and Semantics of Gradability and Comparison*. New York, London: Garland.
- Larson, R. K. 1998. Events and modification in nominals. In D. Strolovitch & A. Lawson (eds.), *Proceedings from SALT VIII*, 145–168. Ithaca: CLC Publication.
- Löbner, S. 1990. *Wahr neben Falsch. Duale Operatoren als die Quantoren natürlicher Sprache*. Tübingen: Niemeyer.
- Löbner, S. 2011. Concept types and determination. *Journal of Semantics* 28. 279–333.
- Löbner, S. 2012a. Sub-compositionality. In W. Hinzen, E. Machery & M. Werning (eds.), *The Oxford Handbook of Compositionality*, 220–241. Oxford: Oxford University Press.
- Löbner, S. 2012b. Frames and Functional Concepts. ms. http://semanticsarchive.net/Archive/j11NGEwO/Loebner_Functional_Concepts_and_Frames.pdf.
- Löbner, S. 2013. *Understanding Semantics*. London, New York: Routledge 2nd edn.
- Löbner, S. 2014. Evidence for frames from natural language. In T. Gamerschlag, D. Gerland, W. Petersen & R. Osswald (eds.), *Frames and Concept Types. Applications in Language and Philosophy*. Studies in Linguistics and Philosophy Vol. 94, 23–68. Dordrecht: Springer.
- Löbner, S. to appear in 2015. The semantics of nominals. In N. Riemer (ed.), *The Routledge Handbook of Semantics*, London, New York: Routledge.
- McNally, L. to appear. Modification. In M. Aloni & P. Dekker (eds.), *The Cambridge Handbook of Semantics*, Cambridge: Cambridge University Press.

- McNally, L. & G. Boleda. 2004. Relational adjectives as properties of kinds. In O. Bonami & P. Cabredo Hofherr (eds.), *Empirical Issues in Formal Syntax and Semantics*, 179–196. Paris: CSSP.
- Morzycki, M. 2005. *Mediated Modification: Functional Structure and the Interpretation of Modifier Position*: University of Massachusetts, Department of Linguistics dissertation.
- Morzycki, M. to appear. *Modification*. Cambridge: Cambridge University Press.
- Naumann, R. 2014. Phase quantification and frame theory. In D. Gerland, C. Horn, A. Latrouite & A. Ortman (eds.), *Meaning and Grammar of Nouns and Verbs*, 237–265. Düsseldorf: Düsseldorf University Press.
- Ó Séaghdha, D. & A. Copestake. 2009. Using lexical and relational similarity to classify semantic relations. In *Proceedings of the 12th Conference of the European Chapter of the ACL*, 621–629.
- Petersen, W. 2007. Representation of Concepts as Frames. In J. Skilters, F. Toccafondi & G. Stemberger (eds.), *Complex Cognition and Qualitative Science*, 151–170. University of Latvia.
- Petersen, W., J. Fleischhauer, H. Beseoglu & P. Bücken. 2008. A frame-based analysis of synaesthetic metaphors. In *The Baltic International Yearbook of Cognition, Logic, and Communication*, vol. 3, New Prairie Press.
- Petersen, W. & T. Osswald. 2012. A formal interpretation of concept types and type shifts. In K. Kosecki & J. Badio (eds.), *Cognitive Processes in Language*, 183–191. Frankfurt: Peter Lang.
- Petersen, W. & T. Osswald. 2014. Concept composition in frames: Focusing on genitive constructions. In T. Gamerschlag, D. Gerland, R. Osswald & W. Petersen (eds.), *Frames and Concept Types. Applications in Language and Philosophy*, 243–266. Dordrecht: Springer.
- Pustejovsky, J. 1995. *The Generative Lexicon*. Cambridge MA: MIT Press.
- Sassoon, G. W. 2013. *Vagueness, Gradability, and Typicality*. Leiden: Brill.
- Schlücker, B. 2014. *Grammatik im Lexikon. Adjektiv-Nomen-Verbindungen im Deutschen*. Berlin: de Gruyter.
- Schulzek, D. 2014. A frame approach to metonymical processes in some common types of German word formation. In T. Gamerschlag, D. Gerland, R. Osswald & W. Petersen (eds.), *Frames and Concept Types. Applications in Language and Philosophy*, 221–242. Dordrecht: Springer.
- Stassen, L. 2003. *Intransitive Predication*. Oxford: Oxford University Press.
- Svenonius, P. 2008. The position of adjectives and other phrasal modifiers in the decomposition of DP. In L. McNally & C. Kennedy (eds.), *Adjectives and Adverbs: Syntax, Semantics, and Discourse*, 16–42. Oxford: Oxford University Press.
- Trost, I. 2006. *Das deutsche Adjektiv*. Hamburg: Buske.
- Vecchi, E. M., R. Zamparelli & M. Baroni. 2013. Studying the recursive behaviour of adjectival modification with compositional distributional semantics. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP 2013)*, 141–151. Seattle.

