G$_h$oSt-NN: A Representative Gold Standard of German Noun-Noun Compounds

Sabine Schulte im Walde, Anna Hätty, Stefan Bott, Nana Khvtisavrishvili
Institute for Natural Language Processing
University of Stuttgart, Germany
{*schulte, anna.haetty, stefan.bott, nana.khvtisavrishvili*}*@ims.uni-stuttgart.de*

Abstract

This paper presents a novel gold standard of German noun-noun compounds *G$_h$ost-NN* including 868 compounds annotated with *corpus frequencies* of the compounds and their constituents, *productivity* and *ambiguity* of the constituents, *semantic relations* between the constituents, and *compositionality ratings* of compound–constituent pairs. Moreover, a subset of the compounds containing 180 compounds is balanced for the productivity of the modifiers (distinguishing low/mid/high productivity) and the ambiguity of the heads (distinguishing between heads with 1, 2 and >2 senses).

Keywords: noun–noun compounds, compositionality ratings, semantic relations

1. Introduction

Compounds are combinations of two or more simplex words. They have been a recurrent focus of attention within theoretical, cognitive, and in the last decade also within computational linguistics. Our focus of interest is on German noun-noun compounds,[1] such as *Ahornblatt* ‘maple leaf’, *Feuerwerk* ‘fireworks’, and *Obstkuchen* ‘fruit cake’, where both the grammatical head (in German, this is the rightmost constituent) and the modifier are nouns. More specifically, we are interested in the degrees of compositionality of German noun-noun compounds, i.e., the semantic relatedness between the meaning of a compound (e.g., *Feuerwerk*) and the meanings of its constituents (e.g., *Feuer* ‘fire’ and *Werk* ‘opus’).

The compositionality of noun-noun compounds has attracted much research across languages over the years. From a psycholinguistic point of view, researchers are interested in finding out how compound words are cognitively processed and represented in the mental lexicon. There is an ongoing debate about whether morphologically complex words are stored and processed as single units (*full listing approach* (Butterworth, 1983)), whether they are *decomposed* into their morphemes (Taft and Forster, 1975; Taft, 2004), or whether they can be accessed both ways: as whole forms and componentially (*dual route models*, cf. Caramazza et al. (1988), Baayen and Schreuder (1999)).

From a computational point of view, addressing the compositionality of noun compounds (and multi-word expressions in more general) is a crucial ingredient for lexicography and NLP applications, to know whether the expression should be treated as a whole, or through its constituents, and what the expression means. For example, studies such as Cholakov and Kordoni (2014); Weller et al. (2014); Cap et al. (2015); and Salehi et al. (2015) have integrated the prediction of compositionality into Statistical Machine Translation. Accordingly, research across languages has aimed for predicting the compositionality of noun compounds automatically. For example, Reddy et al. (2011) predicted the compositionality of 90 English noun–noun compounds via distributional information. Similarly, Schulte im Walde et al. (2013) assessed various types of distributional features to predict the compositionality of 244 German noun–noun compounds. Salehi and Cook (2013) and Salehi et al. (2014) explored multi-lingual dictionaries and distributional evidence to predict the compositionality of German and English noun–noun compounds.

Evaluating predictions of compositionality requires a gold standard of compositionality ratings, if the evaluation is not extrinsic. So in parallel to the emergence of computational systems predicting compositionality, there has also been an increase of gold standards to evaluate the predictions. Regarding noun compounds, Reddy et al. (2011) used heuristics about hypernymy and definitions in WordNet to induce 90 English noun–noun compounds. Schulte im Walde et al. (2013) relied on an existing selection of noun compounds (von der Heide and Borgwaldt, 2009) and used a subset of concrete two-part noun–noun compounds. The work by Salehi et al. used both those datasets.

This paper presents a novel gold standard for the compositionality of German noun-noun compounds. In the next Section 2. we outline the desired properties of the gold standard. Sections 3. and 4. describe in detail how we created the gold standard, and what its resulting properties are.

2. Desired Properties of the Gold Standard

In previous work (Schulte im Walde et al., 2013), we used a gold standard of German noun-noun compositionality ratings that was based on a selection of noun compounds by von der Heide and Borgwaldt (2009). The original target set contained 450 concrete, depictable German noun compounds, with judgements on the compositionality of all compound–constituent pairs. From the compound set by von der Heide and Borgwaldt, we disregarded compounds with more than two constituents as well as compounds where the modifiers were not nouns. Our final set comprised a subset of their compounds including 244 two-part noun-noun compounds.

[1] See Fleischer and Barz (2012) for a detailed overview and Klos (2011) for a recent detailed exploration.

What is the motiviation for creating a novel gold standard for the compositionality of German noun–noun compounds? We were interested in exploring factors that have been found to influence the cognitive processing and representation of compounds, such as

- *frequency-based factors*, i.e., the frequencies of the compounds and their constituents (e.g., van Jaarsveld and Rattink (1988), Janssen et al. (2008));
- the *productivity (morphological family size)*, i.e., the number of compounds that share a constituent (de Jong et al., 2002); and
- semantic variables as the *relationship between compound modifier and head*: a teapot is a pot FOR tea, and a snowball is a ball MADE OF snow (Gagné and Spalding, 2009).
- In addition, we were interested in the effect of *ambiguity* (of both the modifiers and the heads) regarding the compositionality of the compounds.

Consequently, we created a gold standard with a focus on two-part noun-noun compounds including

- compounds and constituents from various frequency ranges;
- compounds and constituents from various productivity ranges;
- compounds and constituents with various numbers of senses; and
- compounds with various semantic relations.

Optimally, the compound targets in the gold standard should be balanced according to all of the above criteria, to include a similar number of compounds and constituents across the conditions. The following section will describe details of the creation process, and to what extent we achieved such a balance.

3. Creation of the Gold Standard

We rely on one of the currently largest corpora for German to induce our new gold standard of German noun–noun compounds *$G_h$ost-NN*: the web corpus *DECOW14AX*[2] (Schäfer and Bildhauer, 2012; Schäfer, 2015), henceforth: *decow*, containing 11.7 billion words. The creation pipeline to acquire a balanced gold standard from the web corpus includes the following steps:

1. corpus-based induction of a noun-noun compound candidate list;
2. addition of empirical properties to the compound candidates;
3. random but balanced selection of a core set of noun-noun compounds;
4. systematic extension of the core set to the full gold standard; and
5. annotation of the gold standard.

3.1. Corpus-based induction of candidate list

Relying on the *decow* corpus, we first extracted all words identified as common nouns by the *Tree Tagger* (Schmid, 1994), plus their lemmas. The noun lemmas were counted, resulting in a total of 365,786 lemma types and their corpus frequencies.

As we wanted to focus on two-part noun-noun compounds only, we applied the morphological analyser *SMOR* (Faaß et al., 2010) to the set of corpus lemmas, providing us potentially ambiguous morphological analyses for all 365,786 noun lemmas. From these, we extracted only those lemmas where SMOR predicted an analysis with exactly two nominal constituents. This set of two-part noun-noun compounds contained 154,960 compound candidate types for our new gold standard.

3.2. Enrichment of empirical properties

The complete set of 154,960 N-N candidates was enriched with empirical properties relevant for the gold standard:

- *corpus frequencies* of the compounds and the constituents (i.e., modifiers and heads), relying on *decow*;
- *productivity* of the constituents (modifiers and heads), i.e., how many compound types contained a specific modifier or head constituent;
- *number of senses* of the constituents (modifiers and heads) and the compounds, relying on GermaNet (Hamp and Feldweg, 1997; Kunze, 2000).

Overall, the 154,960 compound candiates included 7,061 different modifiers and 6,903 different heads.

Figure 5 shows the productivity of the constituents, by plotting the logarithm of the productivity against the number of constituents. For example, 332/361 modifiers/heads appeared in more than 100 different compounds; in contrast, there are approx. 1,388/1,434 modifiers/heads that appeared in only one compound.

Figure 2 shows the number of senses of the modifier and head constituents, as a proportion of those constituent types that were included in GermaNet. Overall, GermaNet covered 26,444 of our 154,960 compound candidate types (17%); 6,683 of the modifier types (95%) and 6,550 of the head types (95%). A large proportion of compound types (97%) has only one sense in GermaNet. In comparison, 65% and 62% of the modifiers/heads have only one sense in GermaNet; 23%/24% have two senses, and below 10% have more than two senses.

3.3. Random but balanced compound selection

From the set of compound candidates we wanted to extract a random subset that at the same time was balanced across frequency, productivity and ambiguity ranges of the compounds and their constituents. Since defining and combining several ranges for each of the three criteria and for compounds as well as constituents would have led to an explosion of factors to be taken into account, we focused on two main criteria instead:

[2] http://corporafromtheweb.org/decow14/

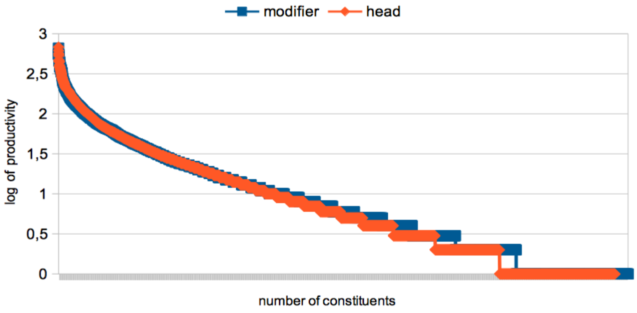

Figure 1: Productivity of constituents.

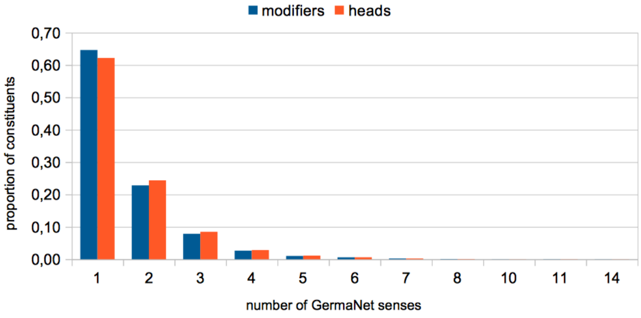

Figure 2: Ambiguity of constituents.

1. *productivity of the modifiers*: We calculated the tertiles to identify modifiers with low/mid/high productivity.
2. *ambiguity of the heads*: We distinguished between heads with 1, 2 and >2 senses.

The total of 9 criteria combinations is listed here:

modifier prod.	low			mid			high		
head senses	1	2	>2	1	2	>2	1	2	>2
modes	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

Table 1: Combined ranges for random extraction.

For each of the 9 categories, we randomly selected 20 noun-noun compounds from our candidate set, disregarding compounds with a corpus-frequency < 2,000, and disregarding compounds containing modifiers or heads with a corpus-frequency < 100. For reasons which will become clear in the following subsection, we also created a subset of 5 noun-noun compounds for each criteria combination, by randomly selecting 5 out of the 20 selected compounds in each mode.

3.4. Systematic extension of core set

We systematically extended the set of 20×9 randomly selected compounds by adding all compounds from the original set of compound candidates (cf. Section 3.1.) with either the same modifier or the same head as any of the selected compounds. Taking *Haarpracht* as an example (the modifier is *Haar* 'hair', the head is *Pracht* 'glory'), we added *Haarwäsche, Haarkleid, Haarpflege, etc.* as well as *Blütenpracht, Farbenpracht, etc.*[3] The total set of target compounds after the systematic extension contained > 5,000 types; relying only on the set of 5×9 randomly selected compounds, it contained 1,208 types. The latter selection appeared as a reasonable size for a gold standard to be annotated for semantic criteria (see below).

While the extension procedure destroyed the coherent balance of criteria that underlied our random extraction described in the previous section, it ensured a variety of compounds with either the same modifiers or the same heads.

3.5. Annotation of gold standard

For computational experiments, researchers can either use the well-balanced set of 20×9=180 compounds without much overlap in modifiers or heads, or the complete G_h*ost-NN* containing 868 compounds out of the larger, less-balanced set of 1,208 compounds.[4] These two sets of compounds were annotated with two kinds of semantic information, (i) the *semantic relations* between the modifiers and the heads, and (ii) *compositionality ratings*.

3.5.1. Semantic relation annotation

In previous work, different kinds of annotation schemes have been used for compound relation annotation. Girju et al. (2005) annotated 282 English two-part noun compounds and 244 English three-part noun compounds with a list of eight prepositional paraphrases previously proposed by Lauer (1995), and also with a set of 35 semantic relations introduced by Moldovan and Girju (2003). Ó Séaghdha (2007) relied on a set of nine semantic relations suggested by Levi (1978), and designed and evaluated a set of relations that took over four of Levi's relations (`BE, HAVE, IN, ABOUT`) and added two relations refering to event participants (`ACTOR, INST(rument)`) that replaced the relations `MAKE, CAUSE, FOR, FROM, USE`. The relation `LEX` refers to lexicalised compounds where no relation can be assigned. A set of 1,443 English noun compounds was annotated with his modified relation set. Dima et al. (2014) worked on German noun compounds and suggested to combine paraphrase- and property-based relation annotation.

We decided to apply the relation set suggested by Ó Séaghdha (2007) to our German noun compounds, for two reasons: (i) He had evaluated his annotation relations and annotation scheme, and (ii) his dataset had a similar size as ours, so we could aim for comparing results across languages. The three authors of this paper who are native speakers of German annotated the compounds with his semantic relations. We used three rounds for the annotation, with discussions in between. If disagreement could not be resolved in the discussions, we disregarded the respective compounds. In the end, we accepted 868 from the 1,208 compounds as gold standard target compounds. These compounds were annotated with the same relation by all three annotators. The distribution of the compounds over the semantic relations is shown in Figure 3.

[3] The translations of the example compounds are *hair washing, hair dress, hair care, floral glory, colour glory*, respectively.

[4] From the enlarged set of 1,208 compounds, the final dataset contains only 868 instances, to ensure a reliable agreement on relation annotation, see Section 3.5.1. below.

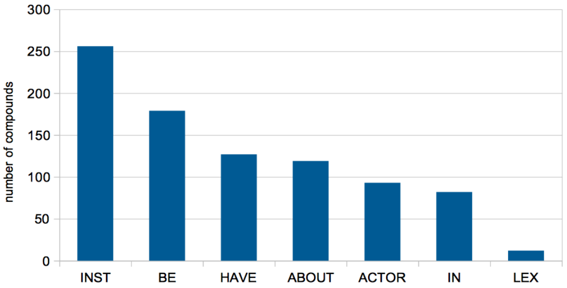

Figure 3: Semantic relations in compounds.

3.5.2. Compositionality ratings

We set up two alternative ways to collect compositionality ratings for the compound–constituent pairs in G_h*ost-NN*:

(1) Expert Ratings On the one hand, 8 native speakers of German annotated all 868 gold-standard compounds with compositionality ratings on a scale from 1 (definitely semantically opaque) to 6 (definitely semantically transparent). Another 5 native speakers provided additional annotation for our small core subset of 5×9=180 compounds on the same scale. All 13 annotators were linguists or computational linguists and did not include any of the authors. Since annotators were allowed to omit judgements if they did not know a compound or a constituent, the actual number of ratings per compound–constituent pair was between 5 and 13.

(2) AMT Ratings We started a collection of compositionality ratings for all 868 gold-standard compounds via Amazon Mechanical Turk[5] (AMT). We randomly distributed the compounds over 40 batches, with 21–22 compounds each, in random order. In order to control for spammers, we also included four German fake compound nouns into each of the batches, in random positions of the lists. If participants did not recognise the fake words, all of their ratings were rejected. By February 2016, we collected between 7 and 15 ratings per compound–modifier and compound–head pair, for 168 of our 868 compounds. The collection is being continued until we have a similar number of ratings for the compound–constituent pairs of all 868 compounds.

4. Properties of the Gold Standard

The final part of this paper aims to illustrate that the new gold standard of German noun-noun compounds covers the ranges of its properties (frequency ranges, degrees of ambiguity, relation types, degrees of compositionality ratings) well, even though we could not balance the extraction according to all these properties.

Figure 4 shows the log frequencies of the compounds, the modifiers and the heads, for all 868 compounds and sorted by the log frequency of the compounds. We can see that neither the modifier nor the head frequencies show a strong bias towards low or high values. The majority of nouns have a log frequency between 5 and 7 (corresponding to approx. 2,000 and 10 million token occurrences). Moreover, there is no tight correlation[6] between compound and modifier frequencies ($\rho = 0.2345, p < 0.001$) and compound and head frequencies ($\rho = 0.1451, p < 0.001$).

Figure 5 compares the log productivity scores of the modifiers and heads within the same compounds. While most compounds seem to include modifiers and heads with log frequencies between 1,5 and 2,8 (roughly corresponding to productivity scores between 30 and 650), other ranges are also covered sufficiently. The correlation between the modifier and head productivity scores confirms their independencies: $\rho = 0.1271, p < 0.001$.

Figure 6 provides the same information regarding the ambiguity of modifiers and heads within compounds. We can see that we cover most combinations of modifiers and heads with ambiguities between 1 (monosemous) and 7. Again, their is no correlation ($\rho = 0.0193$, $p = 0.2840$).

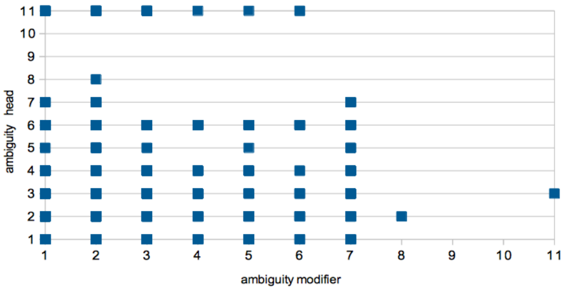

Figure 6: Relation between modifier and head ambiguity.

Figure 7 plots the log frequencies of the compounds (ordered by frequency values of individual compounds) for each semantic relation type. We can see that all relations apply across frequency ranges, i.e., there is no strong bias of specific relations applying only to low vs. mid vs. high frequency compounds.

Similarly, Figure 8 shows that the compositionality ratings are also not due to specific types of semantic relations between modifiers and heads: Each relation type includes compounds with various degrees of compositionality regarding compound–modifier pairs (upper part of the plot) and compound–head pairs (lower part of the plot).

Finally, Figure 9 shows that the productivity of the modifiers/heads (blue dots) is not correlated with specifically low or high compositionality ratings of the compound–modifier and compound–head pairs (red dots), respectively. The correlation between productivity and compositionality is $\rho = -0.0239$ ($p = 0.2421$) for modifiers and $\rho = -0.2043, p < 0.001$ for heads. So there is a very (!) weak negative correlation between the productivity of a compound head and the compositionality of the compound–head pair: the higher the productivity, the lower the compositionality.

5. Availability of the Gold Standard

We provide three resources based on the research in this paper. They are freely available for education, research and other non-commercial purposes:

[5] www.mturk.com

[6] We rely on the Spearman rank-order correlation coefficient ρ (Siegel and Castellan, 1988) to calculate correlations.

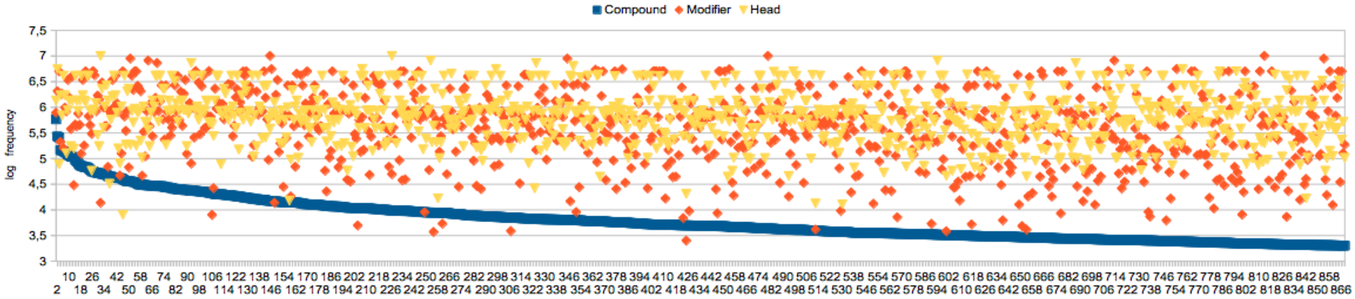

Figure 4: Log frequencies of compounds, modifiers and heads, sorted by compound frequency.

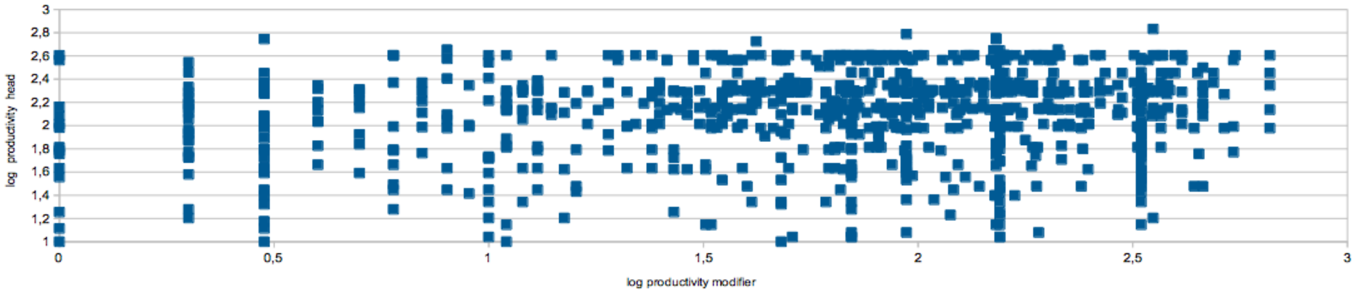

Figure 5: Relation between modifier and head productivity.

1. the set of 154,960 noun-noun candidate compounds and their constituents (cf. Section 3.1.), accompanied by corpus frequency, productivity and degree of ambiguity (cf. Section 3.2.);

2. the final gold standard G_h*ost-NN* of 868 (out of 1,208) noun-noun compounds and their constituents, accompanied by corpus frequency, productivity, ambiguity, and annotated with semantic relations and compositionality ratings (cf. Section 3.5.); and

3. the carefully balanced G_h*ost-NN* subsets of 20×9 and 5×9 compounds and their constituents, categorised according to our 9 criteria combinations for modifier productivity and head ambiguity (cf. Section 3.3.).

For computational experiments, researchers can either use the well-balanced set of 20×9=180 compounds without much overlap in modifiers or heads, or a larger, but less-balanced set of 868 compounds. The datasets are available from `http://www.ims.uni-stuttgart.de/data/ghost-nn`.
Table 2 provides some example compounds and their properties. They were chosen to illustrate the variety of property combinations across items, while at the same time they include compounds with common modifiers or heads.

6. Acknowledgements

The research presented in this paper was funded by the DFG Heisenberg Fellowship SCHU 2580/1 (Sabine Schulte im Walde), the DFG Research Grant SCHU 2580/2 *"Distributional Approaches to Semantic Relatedness"* (Stefan Bott), and the DFG Collaborative Research Center SFB 732 (Anna Hätty, Nana Khvtisavrishvili).

7. Bibliographical References

Baayen, H. and Schreuder, R. (1999). War and Peace: Morphemes and Full Forms in a Noninteractive Activation Parallel Dual–Route Model. *Brain and Language*, 68:27–32.

Butterworth, B. (1983). Lexical Representation. In Brian Butterworth, editor, *Language Production*, pages 257–294. Academic Press, London.

Cap, F., Nirmal, M., Weller, M., and Schulte im Walde, S. (2015). How to Account for Idiomatic German Support Verb Constructions in Statistical Machine Translation. In *Proceedings of the 11th Workshop on Multiword Expressions*, pages 19–28, Denver, Colorado, USA.

Caramazza, A., Laudanna, A., and Romani, C. (1988). Lexical Access and Inflectional Morphology. *Cognition*, 28:297–332.

Cholakov, K. and Kordoni, V. (2014). Better Statistical Machine Translation through Linguistic Treatment of Phrasal Verbs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 196–201, Doha, Qatar.

de Jong, N. H., Feldman, L. B., Schreuder, R., Pastizzo, M., and Baayen, H. R. (2002). The Processing and Representation of Dutch and English Compounds: Peripheral Morphological and Central Orthographic Effects. *Brain and Language*, 81:555–567.

Dima, C., Henrich, V., Hinrichs, E., and Hoppermann, C. (2014). How to Tell a Schneemann from a Milchmann: An Annotation Scheme for Compound-Internal Relations. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*, pages 1194–1201, Reykjavik, Iceland.

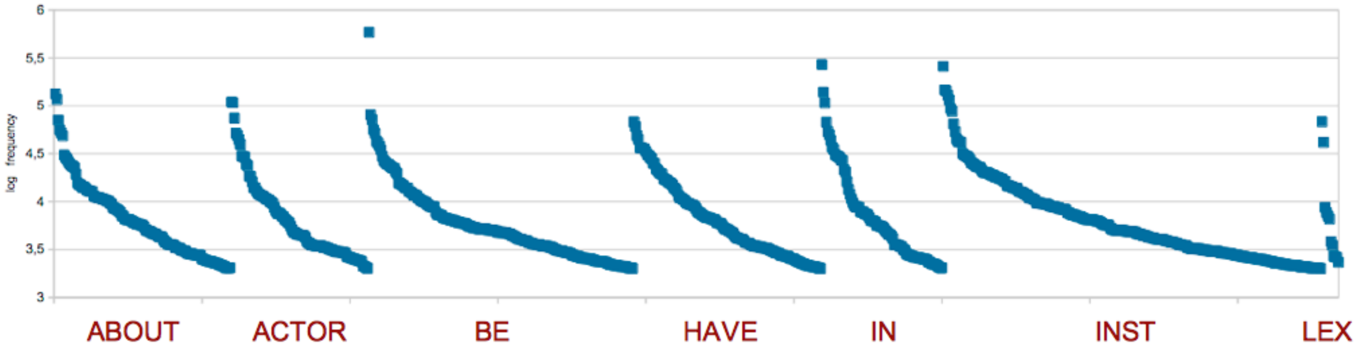

Figure 7: Log frequencies of compounds across relations types.

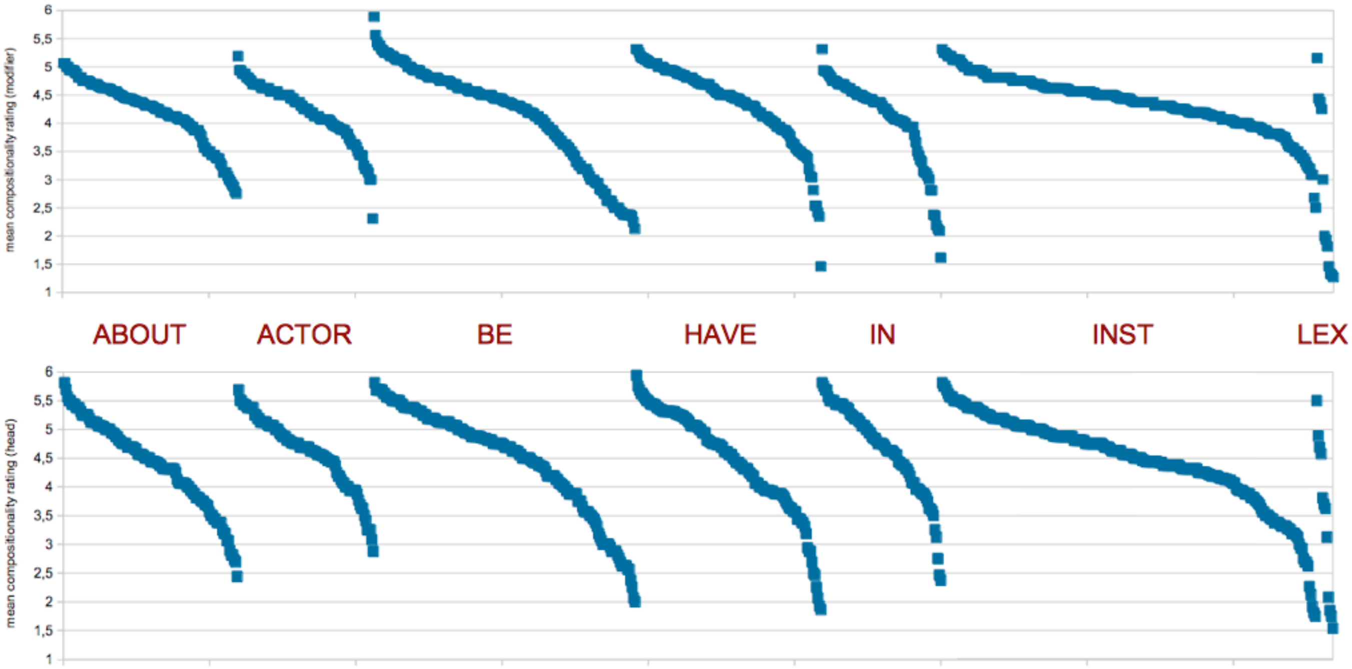

Figure 8: Compositionality ratings across relation types.

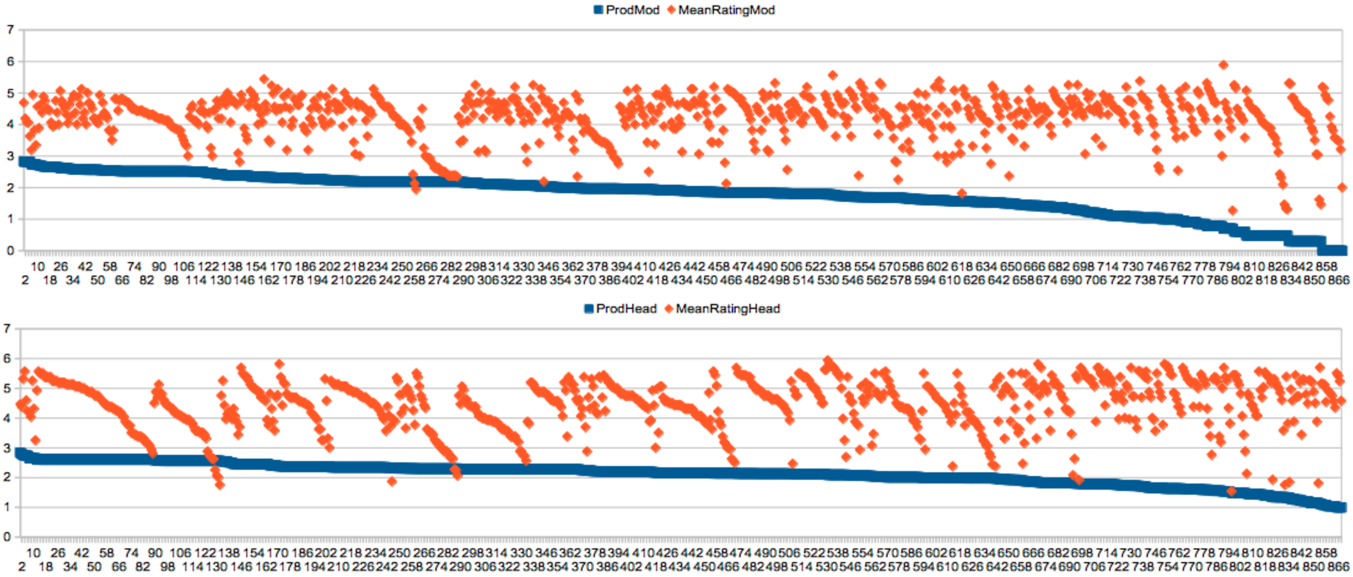

Figure 9: Productivity and compositionality ratings of modifiers and heads.

Nouns						*Frequencies*			*Productivities*		*Ambiguities*		*Relation*	*Ratings*	
Compound		*Modifier*		*Head*		*Compound*	*Modifier*	*Head*	*Modifier*	*Head*	*Modifier*	*Head*		*Modifier*	*Head*
Stadthotel	*city hotel*	Stadt	*city*	Hotel	*hotel*	3,405	4,053,206	1,199,856	543	59	1	1	IN	3.35	5.35
Stadtrand	*suburb*	Stadt	*city*	Rand	*border*	25,099	4,053,206	523,473	543	98	1	2	HAVE	4.94	4.25
Stadtwerk	*public services*	Stadt	*city*	Werk	*plant*	107,754	4,053,206	1,354,148	543	366	1	6	ACTOR	3.81	3.69
Sonnenenergie	*solar energy*	Sonne	*sun*	Energie	*energy*	25,398	832,636	1,191,333	155	30	3	2	INST	4.58	5.44
Sonnenkönig	*Sun King*	Sonne	*sun*	König	*king*	2,680	832,636	494,221	155	109	3	3	LEX	1.94	5.50
Sonnenmasse	*sun mass*	Sonne	*sun*	Masse	*mass*	3,433	832,636	468,284	155	108	3	3	HAVE	4.56	4.75
Sonnenscheibe	*solar disc*	Sonne	*sun*	Scheibe	*slice*	3,155	832,636	364,567	155	96	3	4	BE	4.56	3.75
Sonnenseite	*sunny side*	Sonne	*sun*	Seite	*side*	7,279	832,636	5,508,445	155	256	3	6	IN	4.00	4.31
Sonnenstrahl	*sunbeam*	Sonne	*sun*	Strahl	*beam*	44,612	832,636	32,182	155	27	3	3	HAVE	5.13	4.69
Sonnenuhr	*sundial*	Sonne	*sun*	Uhr	*clock*	8,407	832,636	4,507,590	155	63	3	2	INST	3.75	5.31
Jeanshose	*jeans*	Jeans	*jeans*	Hose	*trousers*	2,971	66,789	273,665	19	61	1	1	BE	5.25	5.44
Latzhose	*overall*	Latz	*bib*	Hose	*trousers*	3,296	5,324	273,665	1	61	2	1	HAVE	3.54	5.23
Strumpfhose	*tights*	Strumpf	*stockings*	Hose	*trousers*	20,535	26,331	273,665	13	61	1	1	BE	4.35	4.42
Kirchspiel	*parish*	Kirche	*church*	Spiel	*game*	6,583	1,761,187	4,122,168	319	403	3	6	LEX	4.44	3.13
Machtspiel	*power game*	Macht	*power*	Spiel	*game*	4,408	806,162	4,122,168	169	403	2	6	ABOUT	4.63	3.44
Ritterspiel	*knights' tournament*	Ritter	*knight*	Spiel	*game*	2,365	115,484	4,122,168	47	403	1	6	ACTOR	3.94	4.75
Testspiel	*tryout*	Test	*test*	Spiel	*game*	37,800	660,169	4,122,168	100	403	3	6	BE	4.25	5.19
Trauerspiel	*fiasco*	Trauer	*grief*	Spiel	*game*	10,763	134,379	4,122,168	77	403	3	6	ABOUT	3.06	2.81
Windspiel	*wind chimes*	Wind	*wind*	Spiel	*game*	2,284	551,317	4,122,168	88	403	3	6	INST	4.31	2.94
Winterspiel	*winter games*	Winter	*winter*	Spiel	*game*	16,067	721,552	4,122,168	207	403	1	6	IN	4.43	5.14
Würfelspiel	*game of dice*	Würfel	*dice*	Spiel	*game*	4,408	80,371	4,122,168	14	403	2	6	INST	4.94	5.56
Bergkette	*mountain chain*	Berg	*mountain*	Kette	*chain*	8,799	564,178	207,479	205	139	2	4	BE	5.13	2.56
Halskette	*necklace*	Hals	*neck*	Kette	*chain*	8,707	271,703	207,479	39	139	3	4	IN	3.94	5.44
Handelskette	*trade chain*	Handel	*trade*	Kette	*chain*	6,509	428,611	207,479	240	139	1	4	INST	4.75	3.38
Hotelkette	*hotel chain*	Hotel	*hotel*	Kette	*chain*	6,410	1,199,856	207,479	134	139	1	4	BE	5.00	3.13
Menschenkette	*human chain*	Mensch	*human*	Kette	*chain*	6,383	8,884,087	207,479	191	139	1	4	BE	4.94	3.75
Produktionskette	*production chain*	Produktion	*production*	Kette	*chain*	2,738	579,419	207,479	244	139	2	4	HAVE	4.69	3.19
Schneekette	*snow chains*	Schnee	*snow*	Kette	*chain*	5,167	324,839	207,479	95	139	1	4	INST	4.19	4.21
Zeichenkette	*string*	Zeichen	*character*	Kette	*chain*	8,836	749,903	207,479	62	139	3	4	BE	4.34	2.95

Table 2: Example compounds in $G_h ost\text{-}NN$ and their properties.

Faaß, G., Heid, U., and Schmid, H. (2010). Design and Application of a Gold Standard for Morphological Analysis: SMOR in Validation. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, pages 803–810, Valletta, Malta.

Fleischer, W. and Barz, I. (2012). *Wortbildung der deutschen Gegenwartssprache*. de Gruyter, Berlin.

Gagné, C. L. and Spalding, T. L. (2009). Constituent Integration during the Processing of Compound Words: Does it involve the Use of Relational Structures? *Journal of Memory and Language*, 60:20–35.

Girju, R., Moldovan, D., Tatu, M., and Antohe, D. (2005). On the Semantics of Noun Compounds. *Journal of Computer Speech and Language*, 19(4):479–496. Special Issue on Multiword Expressions.

Hamp, B. and Feldweg, H. (1997). GermaNet – A Lexical-Semantic Net for German. In *Proceedings of the ACL Workshop on Automatic Information Extraction and Building Lexical Semantic Resources for NLP Applications*, pages 9–15, Madrid, Spain.

Janssen, N., Bi, Y., and Caramazza, A. (2008). A Tale of Two Frequencies: Determining the Speed of Lexical Access for Mandarin Chinese and English Compounds. *Language and Cognitive Processes*, 23:1191–1223.

Klos, V. (2011). *Komposition und Kompositionalität*. Number 292 in Germanistische Linguistik. Walter de Gruyter, Berlin.

Kunze, C. (2000). Extension and Use of GermaNet, a Lexical-Semantic Database. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation*, pages 999–1002, Athens, Greece.

Lauer, M. (1995). *Designing Statistical Language Learners: Experiments on Noun Compounds*. Ph.D. thesis, Macquarie University, Australia.

Levi, J. N. (1978). *The Syntax and Semantics of Complex Nominals*. Academic Press, London.

Moldovan, D. and Girju, R. (2003). Tutorial on *Knowledge Discovery from Text*. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, Sapporo, Japan.

Ó Séaghdha, D. (2007). Designing and Evaluating a Semantic Annotation Scheme for Compound Nouns. In *Proceedings of Corpus Linguistics*, Birmingham, UK.

Reddy, S., McCarthy, D., and Manandhar, S. (2011). An Empirical Study on Compositionality in Compound Nouns. In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, pages 210–218, Chiang Mai, Thailand.

Salehi, B. and Cook, P. (2013). Predicting the Compositionality of Multiword Expressions Using Translations in Multiple Languages. In *Proceedings of the 2nd Joint Conference on Lexical and Computational Semantics*, pages 266–275, Atlanta, GA.

Salehi, B., Cook, P., and Baldwin, T. (2014). Using Distributional Similarity of Multi-way Translations to Predict Multiword Expression Compositionality. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 472–481, Gothenburg, Sweden.

Salehi, B., Mathur, N., Cook, P., and Baldwin, T. (2015). The Impact of Multiword Expression Compositionality on Machine Translation Evaluation. In *Proceedings of the 11th Workshop on Multiword Expressions*, pages 54–59, Denver, Colorado, USA.

Schäfer, R. and Bildhauer, F. (2012). Building Large Corpora from the Web Using a New Efficient Tool Chain. In *Proceedings of the 8th International Conference on Language Resources and Evaluation*, pages 486–493, Istanbul, Turkey.

Schäfer, R. (2015). Processing and Querying Large Web Corpora with the COW14 Architecture. In *Proceedings of the 3rd Workshop on Challenges in the Management of Large Corpora*, pages 28–34, Mannheim, Germany.

Schmid, H. (1994). Probabilistic Part-of-Speech Tagging using Decision Trees. In *Proceedings of the 1st International Conference on New Methods in Language Processing*.

Schulte im Walde, S., Müller, S., and Roller, S. (2013). Exploring Vector Space Models to Predict the Compositionality of German Noun-Noun Compounds. In *Proceedings of the 2nd Joint Conference on Lexical and Computational Semantics*, pages 255–265, Atlanta, GA.

Siegel, S. and Castellan, N. J. (1988). *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, Boston, MA.

Taft, M. and Forster, K. I. (1975). Lexical Storage and Retrieval of Prefixed Words. *Journal of Verbal Learning and Verbal Behavior*, 14:638–648.

Taft, M. (2004). Morphological Decomposition and the Reverse Base Frequency Effect. *The Quarterly Journal of Experimental Psychology*, 57:745–765.

van Jaarsveld, H. J. and Rattink, G. E. (1988). Frequency Effects in the Processing of Lexicalized and Novel Nominal Compounds. *Journal of Psycholinguistic Research*, 17:447–473.

von der Heide, C. and Borgwaldt, S. (2009). Assoziationen zu Unter-, Basis- und Oberbegriffen. Eine explorative Studie. In *Proceedings of the 9th Norddeutsches Linguistisches Kolloquium*, pages 51–74.

Weller, M., Cap, F., Müller, S., Schulte im Walde, S., and Fraser, A. (2014). Distinguishing Degrees of Compositionality in Compound Splitting for Statistical Machine Translation. In *Proceedings of the 1st Workshop on Computational Approaches to Compound Analysis*, pages 81–90, Dublin, Ireland.