Aligning Word Senses in GermaNet and the DWDS Dictionary of the German Language

Verena Henrich, Erhard Hinrichs, Reinhild Barkey
Department of Linguistics
University of Tübingen, Germany
{vhenrich,eh,rbarkey}@sfs.uni-tuebingen.de

Abstract

A comparison and alignment of lexical resources brings about considerable mutual benefits for all resources involved. For all sense distinctions that are completely parallel in two resources, such an alignment provides supporting external evidence for the validity of sense distinction and allows enriching word senses by information contained in the other resource. By contrast, for all non-matching sense distinctions, reason for revisiting and possibly revising the lexical entries in question is provided. The purpose of this paper is to compare the German wordnet GermaNet with the Digital Dictionary of the German Language (DWDS) and to align word senses in the two resources. The paper presents issues that arise in practice when such an alignment is performed and indicates the benefits that both resources will gain.

1 Introduction

It has long been recognized that the identification and differentiation of word senses is one of the harder tasks that lexicographers face. As a result, lexical resources display considerable variation in the number of word senses that lexicographers assign to a given lexical entry in a dictionary. Against this background, lexicographic practice has undertaken considerable efforts to find external knowledge sources that can aid in distinguishing and identifying word senses. The external knowledge sources that are most widely used for this purpose are very large electronic corpora that can be harvested for a given word under lexicographic consideration. Another type of resource that has also been explored as an external reference point is the comparison with another semantic dictionary that has been constructed independently for the same language.

The present paper reports on an ongoing project in which the German wordnet GermaNet (Hamp and Feldweg, 1997; Henrich and Hinrichs, 2010) is compared to the word senses contained in the Digital Dictionary of the German Language (*Digitales Wörterbuch der Deutschen Sprache*[1], DWDS; Klein and Geyken, 2010). Both resources are long-term lexicographic projects aiming at a comprehensive coverage of contemporary standard German in electronic form. What makes a comparison between these resources particularly interesting and useful is the fact that they utilize two different methods for constructing word meanings.

The DWDS is based on the digital versions of three pre-existing dictionaries: the Dictionary of Contemporary German (*Wörterbuch der deutschen Gegenwartssprache*, WDG), the Grimm Dictionaries *Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm* (1DWB) and its revised version (2DWB), as well as the Etymological German Dictionary by Wolfgang Pfeifer (EtymWb). The lexical entries inherited from these dictionaries have been revised and amended by information harvested from large electronic corpora of contemporary German (Didakowski et al., 2012). DWDS lexical entries are structured by the number of senses which may be further differentiated by an enumeration of subsenses. Senses are accompanied by examples harvested from German text corpora or by so-called *competence examples* that are manually constructed.

The conception of word meaning underlying GermaNet adheres to the idea of a network of meaningfully related words and concepts that are interlinked by a set of lexical and conceptual relations and that was first realized in the Princeton WordNet for English (Fellbaum, 1998). The set of lexical and conceptual relations include synonymy, hypernymy/hyponymy, meronymy/holonymy, causation, antonymy, and pertainymy.

The comparison of GermaNet and the DWDS dictionary will focus on the alignment of Germa-

[1] http://www.dwds.de/

Net senses (synsets and lexical units) with the senses and subsenses of DWDS lexical entries. The benefits of this GermaNet-DWDS comparison include the following:

- If the set of sense distinctions match for a given word lemma in both resources, then this provides supporting external evidence for the validity of these sense distinctions.
- If the set of sense distinctions differ between the two resources, then this provides reason for revisiting and possibly revising the lexical entries in question.

Apart from the comparison of word senses, each resource stands to gain from the GermaNet-DWDS mapping in the following ways:

- It becomes possible to implement an intelligent semantic search for the DWDS that provides users not just with the word senses of a given lexical entry but also with lexical information about related words.
- GermaNet synsets and lexical units can be enriched by suitable definitions as well as examples contained in the DWDS.

The purpose of this paper is to present the results of a pilot study that concentrates on a set of issues that arise in practice when such a mapping is performed.

2 Survey of the Overlapping Coverage

The total number of lemmas that have lexical entries in both resources is 48,036[2] (6,211 adjectives, 34,366 nouns, and 7,735 verbs), which covers about 53.5% of all lemmas encoded in GermaNet. At first glance, this overlap might seem low. However, on a closer look, there is an explanation for this which mainly concerns the following three points:

- The history of the two resources causes differences in coverage. The DWDS is based on the digital versions of three pre-existing dictionaries that do not include most recent contemporary language. By contrast, the terms to be included in GermaNet follow frequency lists extracted from large corpora such as newspaper texts and Wikipedia, which also contain recent contemporary language.
- Both resources have different basic decisions on what terms and senses should be included. The perspectives and guidelines that the lexicographers of both resources pursue differ. For example, the resources deviate in the inclusion of regional, obsolete, technical, and colloquial terms as well as most recent contemporary language. This further explains why the coverage of GermaNet and the DWDS differs.
- Since compounding is a highly productive phenomenon of the German language, the question of which compounds to include in a lexical resource is not trivial to answer. There are many newly created compounds that eventually – after some undefined time and depending on the frequency of general usage – might become part of the fundamental vocabulary of the German language. Thus, especially for the coverage of compounds, there is a huge deviation between the two resources.

Since senses in the DWDS might be further differentiated by an enumeration of subsenses, a survey on word senses involves more than one comparison. GermaNet distinguishes 59,495 senses for the 48,036 lemmas that the two resources share. The overall number of 61,053 main sense distinctions in the DWDS is very similar. On the contrary, the number of main senses plus subsenses on the highest level encoded in the DWDS is 74,346, which is more than in GermaNet. This suggests a mapping on the main sense level of the DWDS.

The outcome of this survey proves that there is a considerable overlap of word lemmas with a comparable amount of senses in both resources, which supports the usefulness of conducting a sense alignment.

3 Evaluation of the Sense Alignment

In order to be able to evaluate the alignment on the level of senses and subsenses, the lexical entries for an initial set of 470 randomly selected word lemmas (see Section 4 for the selection process) have been manually analyzed with regard to the appropriateness of matching senses from one resource onto the senses in the other resource. The variability of how good the senses can be matched leads to a division into four classes that are illustrated and described in the following four subsections – in descending order according to their alignment appropriateness.

[2] All numbers are calculated on GermaNet's current release 8.0 as of April 2013 and on the DWDS subset taken from version 0.4.17 and filtered for all lexical entries for lemmas covered by both resources. This filtered subset has been made available to us on August 9, 2013.

Figure 1: Sense mapping using the example of *Pferd* (class 1).

3.1 Class 1: Exact match of main senses

Class 1 represents the ideal case, i.e., senses in GermaNet correspond to main senses in the DWDS. The German noun *Pferd* is a case in point. As illustrated in Figure 1, this lemma has the three distinct senses in both resources representing an animal horse, a gymnastic horse, and a chess knight. All word senses that fall into this class show an identical overlapping lexical coverage and an identical granularity level of sense distinctions. For both GermaNet and the DWDS, this provides mutual supporting evidence for the validity of these sense distinctions.

For GermaNet, the obvious gain for all these senses is an enrichment by suitable definitions and examples contained in the corresponding DWDS senses. For the DWDS, it becomes possible for all these senses that an intelligent semantic search provides users not just with the word senses of a given lexical entry but also with lexical information about related words.

3.2 Class 2: Exact match of subsenses

There are several senses in GermaNet that do not correspond to main senses in the DWDS but which correspond to subsenses in the DWDS. These latter ones are included in class 2. Figure 2 gives an example using the word *Bogen*. In GermaNet, there are two distinct senses representing a violin bow and a bow as a weapon (see the left side of Figure 2). In the DWDS, there is a main sense described as *gebogenes Gerät* 'curved device' which is further differentiated into the two subsenses of a violin bow (*sub 1*) and a bow as a weapon (*sub 2*) – see the two entries denoted by *sub 1* and *sub 2* on the right side of Figure 2.

The overall coverage for these senses is the same. It is only the granularity level of the sense distinctions that differs. The reason for this difference results from different perspectives and guidelines of how to model word senses that the lexicographers of both resources pursue. There is an agreement between lexicographers of both resources that the two senses under consideration should be modeled separately. The question of whether to constitute two separate word senses or two subsenses of a common main sense is bound to the nature of the resources, i.e., GermaNet does not further distinguish word senses into subsenses.

The senses that fall into this class again provide support for the validity of the sense distinctions for both resources. Furthermore, the enrichment of GermaNet senses with definitions and examples as well as the enrichment of DWDS senses with information on related words is equally possible than it is described for class 1.

3.3 Class 3: Partly overlapping coverage and different sense distinctions

Class 3 contains senses for which a straightforward one-to-one mapping is not possible. This includes the following two cases: (i) two separate senses from one resource are jointly represented by only one sense in the other resource and (ii) the core meaning of two senses is the same, but the two senses are still not completely identical in their coverage.

The German noun *Pranke* is a case in point for case (i). The DWDS encodes a sense defined as *Vordertatze, besonders von großen Raubtieren; umgangssprachlich, scherzhaft, übertragen: große, starke Hand* 'forepaw of an animal, especially a predator; colloquial, jokingly, figurative: big, strong hand' (see the right side in Figure 3). In GermaNet, *Pranke* has the two fine-grained senses denoting the paw of an animal and the

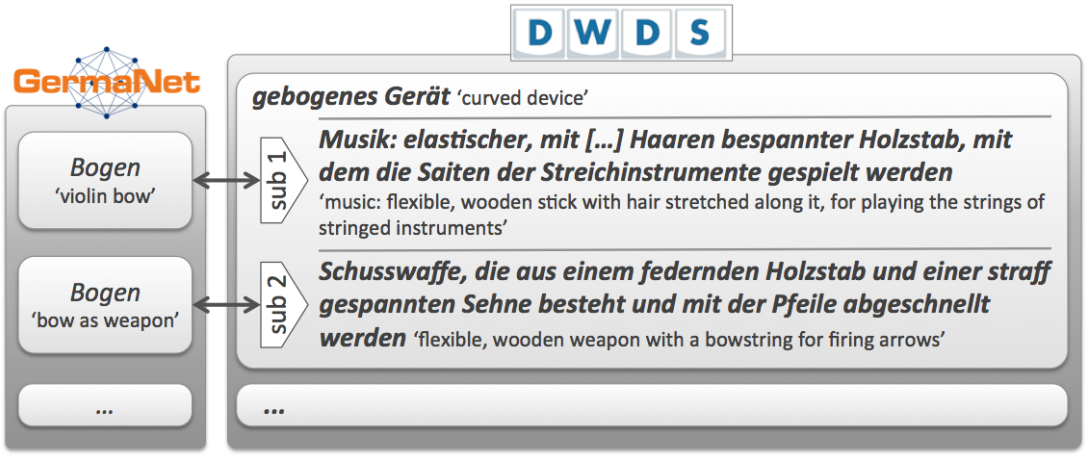

Figure 2: Sense mapping using the example of *Bogen* (class 2).

figurative term for the human hand (see the left side in Figure 3). In this example, both these more specific GermaNet senses are subsumed under one single DWDS sense.

In the second case (ii) that is subsumed by class 3, there is no complete coverage of the meaning of one sense in one resource with the corresponding sense in the other resource. The core sense is mostly identical, but there are meaning aspects that led the lexicographers to decide differently on whether to explicitly encode a separate sense in the dictionary or not.

An example of this type is the German noun *Sturm* 'storm'. Both GermaNet and the DWDS encode a sense referring to the weather phenomenon. Accompanying example sentences of this word sense in the DWDS include instances exemplifying a figurative usage, such as, for example, *ein Sturm der Entrüstung* 'a storm of indignation'. That means, the figurative meaning of *Sturm* is explicitly mentioned in the DWDS weather phenomenon sense – without encoding a separate sense or subsense. By contrast, the figurative meaning of *Sturm* is not present in GermaNet – neither as part of the corresponding weather phenomenon sense nor explicitly as a separate sense.

The phenomena of both cases (i) and (ii) cannot solely be explained by the lexicographic background of the two resources. They rather illustrate different lexicographic perspectives of how to distinguish senses of a word. The question at what point a meaning should be regarded as a distinct sense or subsense to be included in a dictionary is a difficult issue in lexicographic work. Aspects that affect this decision include figurative meaning, technical, colloquial, or regional usage of a term. Both in the paw and in the storm examples, the lexicographers of the two resources have made different decisions with respect to the status of the figurative meaning of a word sense.

As for the benefit from a mapping of senses in this class, it would mean that each example sentence for the DWDS senses in question has to be analyzed individually in order to decide whether it can be assigned to a GermaNet sense. Nonetheless, it is interesting to further analyze these cases since they concern the identification and differentiation of word senses which is one of the harder tasks that lexicographers face.

3.4 Class 4: Distinct coverage

This class comprises lemmas where there is at least one sense or subsense in one resource that does not have a corresponding entry in the other resource. An example of this kind is illustrated in Figure 4 using the example of *Maus*. For this word, GermaNet encodes the two senses of the mouse as an animal and the computer mouse (see the left side of Figure 4). The DWDS also encodes the animal sense of a mouse, but it does not include the computer mouse sense. Instead, the DWDS lists *Mäuse* (plural for *Maus*) in the sense of an informal synonym for money (see the right side of Figure 4).

As illustrated in the mouse example, both resources gain benefit from a sense alignment by mutually providing suggestions of possibly missing senses. In general, with the help of simple word comparisons, it is easy to automatically compile lists of lemmas that serve as candidates to be inserted into a dictionary. By contrast, it is much more difficult to provide (automatic) suggestions of possibly missing senses. In all cases where the sense alignment discovers different

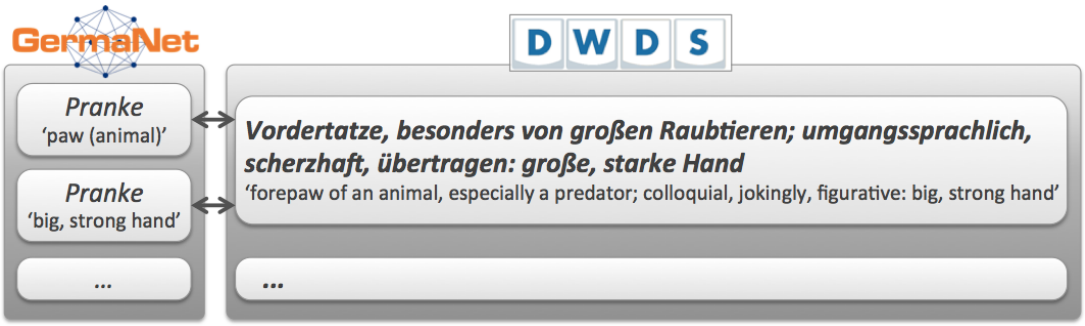

Figure 3: Sense mapping using the example of *Pranke* (class 3).

sets of sense distinctions between GermaNet and the DWDS, this provides reason for revisiting and possibly revising the lexical entries in question.

4 Evaluation Statistics

The selection of the initial set of manually aligned word lemmas is guided by the following criteria:

- The selected words include all three word classes of adjectives, nouns, and verbs.
- In order to ensure a detailed evaluation of lexical items with different degrees of polysemy, the evaluation reports results for five different polysemy classes: words having (i) one sense in GermaNet, (ii) two senses in GermaNet, (iii) three or four senses, (iv) five to ten senses, and (v) more than ten senses in GermaNet.
- The sample as a whole represents a good balance of word classes and number of distinct word senses.

That is, for adjectives and verbs, 35 lemmas were randomly selected for each of the polysemy classes (i) to (v). Since the coverage for nouns is higher compared to the coverage of the other two word classes, 50 nominal lemmas were randomly chosen for each polysemy class. Table 1 shows the total number of word lemmas and corresponding word senses (in parentheses) in each polysemy class for the three word classes[3] that were manually aligned by two experienced lexicographers. Column *All POS* contains the summed numbers for all word classes (i.e., part-of-speech, POS) separately for the polysemy classes.

[3] The information both about the number of distinct word senses as well as about the word category of the lemmas is taken from GermaNet.

Senses	*Adjectives*	*Nouns*	*Verbs*	*All POS*
1	35 (35)	50 (50)	35 (35)	120 (120)
2	35 (70)	50 (100)	35 (70)	120 (240)
3 – 4	35 (114)	50 (161)	35 (112)	120 (387)
5 – 10	8 (51)	50 (282)	35 (209)	93 (542)
> 10	–	3 (36)	14 (192)	17 (228)
Total	113 (270)	203 (629)	154 (618)	470 (1,517)

Table 1: Aligned word lemmas (corresponding word senses in parentheses) and their sense distributions

Note that the number of lemmas for adjectives with three or four senses and for nouns and verbs with more than ten senses is lower than mentioned above. The reason is simply because there are only few lemmas encoded both in GermaNet and the DWDS that fall into these classes, i.e., 8, 3, and 14, respectively. Adjectives with more than ten senses do not exist at all.

Altogether, 470 distinct word lemmas were manually checked by the lexicographers. These lemmas correspond to 1,517 senses (in GermaNet) of which 113 adjectives, 203 nouns, and 154 verbs. That is, the 470 words have an average of 3.2 senses (2.4 for adjectives, 3.1 for nouns, and 4.0 for verbs). With the help of the manual sense alignment, it is possible to classify senses according to their alignment appropriateness, i.e., into classes 1 to 4 described in Sections 3.1-3.4.

Table 2 lists the counts of these 1,517 GermaNet senses classified into the four alignment classes separately for the previously defined polysemy classes (columns). The rightmost column depicts the overall results without classifying words with respect to their number of different senses. The rows show the different alignment classes 1 – 4 separately for each of the three word categories of adjectives, nouns, and verbs. The last row (*All cl.*) sums all aligned senses for each word class per polysemy class. Rows marked with Σ denote results for all word categories.

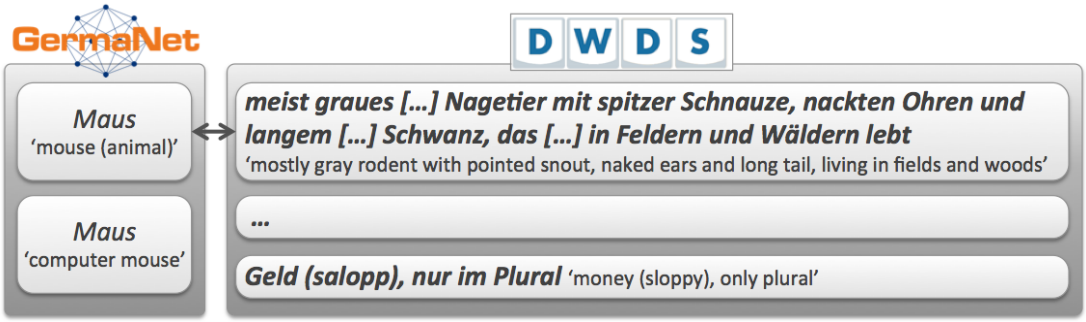

Figure 4: Sense mapping using the example of *Maus* (class 4).

		Senses in GermaNet					
		1	2	3–4	5–10	> 10	Total
Class 1	adj.	35	29	46	16	–	126 (47%)
	nouns	49	64	77	136	9	335 (53%)
	verbs	34	36	51	73	56	250 (40%)
	Σ	118	129	174	225	9	**711 (47%)**
Class 2	adj.	0	5	19	12	–	36 (13%)
	nouns	0	2	11	43	4	60 (10%)
	verbs	0	0	7	48	55	110 (18%)
	Σ	0	7	37	103	4	**206 (14%)**
Class 3	adj.	0	35	38	19	–	92 (34%)
	nouns	1	18	54	58	22	153 (24%)
	verbs	0	32	46	71	71	220 (36%)
	Σ	1	85	138	148	22	**465 (31%)**
Class 4	adj.	0	1	11	4	–	16 (7%)
	nouns	0	16	19	45	1	81 (13%)
	verbs	1	2	8	17	10	38 (6%)
	Σ	1	19	38	66	1	**135 (9%)**
All Cl.	adj.	35	70	114	51	–	270 (100%)
	nouns	50	100	161	282	36	629 (100%)
	verbs	35	70	112	209	192	618 (100%)
	Σ	120	240	387	542	228	1,517 (100%)

Table 2: Sense distribution of the different alignment classes

The numbers in Table 2 count senses rather than lemmas. Note that this implies that senses of a single lemma do not necessarily all have to be classified to the same alignment class but can belong to different classes – what arises quite frequently in practice. An example of this kind is the lemma *Maus* which has already been discussed in Section 3.4 (see Figure 4). The first GermaNet sense depicting the mouse as an animal has a corresponding main sense on the DWDS side; meaning that this sense is counted for alignment class 1. On the contrary, the second GermaNet sense for this lemma, which represents the computer mouse sense, does not have a corresponding match on the DWDS side. Thus, the second sense has to be counted for class 4.

5 Discussion of the Results

To begin with the most prominent and important result, classes 1 (exact match of main senses) and 2 (exact match of subsenses) together arise in 61% of all cases, i.e., 47% and 14%, respectively – see Table 2. This suggests that for three out of five word senses from GermaNet there is a matching sense in the DWDS with which a GermaNet sense can be aligned. This underscores the overall feasibility of a sense alignment between the two lexical resources. The obvious gain for all these senses is the mutual enrichment by sense-specific information – such as suitable definitions, examples, and lexical relations – taken from the matching sense.

Class 1 arises in 47% of all cases and thus much more frequently than all other classes. The fact that matches between GermaNet senses and main senses in the DWDS (class 1) outnumber matches between GermaNet senses and subsenses in the DWDS (class 2) was to be expected. This confirms the conception of word senses on the same granularity level in both resources.

Both classes 3 (partly overlapping coverage and different sense distinctions) and 4 (distinct coverage) reveal differences between GermaNet and the DWDS that prevent a straight forward sense alignment. The explanation why class 3 arises in 31% of all cases, i.e., why there are differences in the distinction of senses, is due to the lexicographic background of the two resources. The lexicographers of GermaNet and the DWDS pursue different perspectives and guidelines of how to model word senses, e.g. with respect to the sense granularity. Thus, from a lexicographer's perspective, it is interesting to analyze these cases since they concern the identification and differentiation of word senses which is one of the harder tasks that lexicographers face. To gain benefit from a mapping of senses in this class, it would mean that all information for a

sense has to be analyzed in order to be individually assigned to a corresponding sense.

Class 4, which indicates a distinct coverage of GermaNet and the DWDS, shows fewest occurrences. In only 9% of all GermaNet senses, there is no corresponding entry in the DWDS. It should be kept in mind that this number only applies to those 48,036 lemmas that are encoded in both resources. For all remaining lemmas, there are no lexical entries in the DWDS at all and thus these word senses would fall into class 4 as well. The evaluation for class 4 is biased towards one direction, i.e., it regards GermaNet senses with missing entries in the DWDS. Since it is also interesting to analyze and compare the other way around where there are DWDS senses lacking matches in GermaNet, these cases have also been recorded during the manual alignment. Altogether, there are 384 word senses (122 adjectival, 104 nominal, and 158 verbal senses) in the DWDS that do not have a corresponding entry in GermaNet. In all cases where the sets of sense distinctions differ between the two resources, this provides reason for revisiting and possibly revising the lexical entries in question. Of course, this also applies to all those word lemmas for which there is a lexical entry in only one of the two resources.

A comparison of the results for the three different word classes and polysemy classes yields the following tendencies:

- Words with only one GermaNet sense almost exclusively fall into class 1 – for all three word classes. This is not surprising since those words usually have one or few senses in the DWDS and thus the probability that the "same" most prominent sense of a word is encoded in both resources is significant.
- More than half of all nouns (53%) fall into class 1 – much fewer nouns (10%) fall into class 2. By contrast, there are only 40% of all verbs in class 1, but proportionally almost twice as many verbs (18%) classified to class 2 compared to nouns. This is especially remarkable for verbs with more than four senses. One reason for this difference is the variety in the granularity level of the sense distinctions in GermaNet and the DWDS which arises more often for verbs than for nouns.
- The deviation between the three word classes for polysemous words, i.e., words with more than one sense in GermaNet, is interesting to observe. Adjectives and verbs show a proportionally larger number of occurrences in class 3 (34% and 36%, respectively) compared to nouns (24%). This means that there are more words with a partly overlapping coverage and different sense distinctions for adjectives and verbs than for nouns, e.g., where two senses from one resource jointly describe one sense of the other resource.
- By contrast, this ratio is reversed for class 4, where there are proportionally nearly twice as many occurrences for nouns (13%) than for adjectives and verbs (7% and 6%, respectively). The explanation for this is that there are more nominal senses that are not encoded in one resource, but more adjectival and verbal senses that encode an overlapping coverage with a different distinction of senses.

All in all, it is worthwhile to perform a complete sense alignment between GermaNet and the DWDS. This will open up a wide range of benefits for both resources, including the harvesting of sense-specific information and the external support of sense distinctions for matching senses as well as indicators for revisiting and possibly revising the lexical entries in question for non-matching senses.

6 Related Work

There has been a considerable body of research for English that investigates the alignment of the Princeton WordNet with Wikipedia (including Ruiz-Casado et al., 2005; Ponzetto and Navigli, 2010; Niemann and Gurevych, 2011), with Wiktionary (including Meyer and Gurevych, 2011), with the Longman Dictionary of Contemporary English and with Roget's thesaurus (Kwong, 1998), with the Hector lexicon (Litkowski, 1999), or with the Oxford Dictionary of English (Navigli, 2006).

Previous work for German has been on the alignment of GermaNet with the German version of Wiktionary (Henrich et al., 2011) and with the German Wikipedia (Henrich et al., 2012).

However, there is no other previous research that tries to align GermaNet to the DWDS.

7 Conclusion and Future Work

This initial pilot study has proven the feasibility of a sense alignment between GermaNet and the DWDS both in term of quantity and appropriateness. We have learned about the differences in the distinction of senses that are due to different perspectives and guidelines of how to model word senses that the lexicographers of both resources pursue. The classification of senses ac-

cording to their appropriateness to be aligned with senses from the other resource allows an individual treatment of different issues and phenomena that arise in practice when an alignment of two resources is performed.

The alignment of GermaNet with the DWDS brings about considerable mutual benefits for both resources. For all sense distinctions that are completely parallel in the two resources, the alignment provides supporting external evidence for the validity of sense distinction and allows enriching word senses by information contained in the other resource. By contrast, for all non-matching sense distinctions, reason for revisiting and possibly revising the lexical entries in question is provided.

The natural next step, which we have already started to work on, is to implement an algorithm that automatically aligns senses from the two resources. This provides a good basis for the lexicographer's work of post-correcting the automatic alignment and revising the senses in both resources, which still remains a complex and substantial task to be performed.

Acknowledgments

We are very grateful to our student assistants Sabrina Galasso and Amit Vrubel, who helped us with the evaluation reported in Sections 4 and 5. Special thanks go to our colleagues in Berlin for making available the DWDS to us.

Financial support was provided by the German Research Foundation (DFG) as part of the Collaborative Research Center 'Emergence of Meaning' (SFB 833), by the German Ministry of Education and Technology (BMBF) as part of the research grant CLARIN-D, and by a research grant from the University of Tübingen.

References

Jörg Didakowski, Lothar Lemnitzer, and Alexander Geyken. 2012. *Automatic example sentence extraction for a contemporary German dictionary*. Proceedings of EURALEX 2012, Oslo, pp. 343-349.

Christiane Fellbaum. (eds.). 1998. WordNet – An Electronic Lexical Database. The MIT Press.

Birgit Hamp and Helmut Feldweg. 1997. *GermaNet – a Lexical-Semantic Net for German*. Proceedings of ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications, Madrid.

Verena Henrich and Erhard Hinrichs. 2010. *GernEdiT – The GermaNet Editing Tool*. Proceedings of the Seventh Conference on International Language Resources and Evaluation, pp. 2228-2235.

Verena Henrich, Erhard Hinrichs, and Klaus Suttner: *Automatically Linking GermaNet to Wikipedia for Harvesting Corpus Examples for GermaNet Senses*. Journal for Language Technology and Computational Linguistics (JLCL), Vol. 27, No. 1, 2012, pp. 1-19.

Verena Henrich, Erhard Hinrichs, and Tatiana Vodolazova. 2011. *Semi-Automatic Extension of GermaNet with Sense Definitions from Wiktionary*. Proceedings of 5th Language & Technology Conference (LTC 2011), Poznań, Poland, 2011, pp. 126-130.

Wolfgang Klein and Alexander Geyken. 2010. *Das Digitale Wörterbuch der Deutschen Sprache (DWDS)*. Heid, Ulrich/Schierholz, Stefan/Schweickard, Wolfgang/Wiegand, Herbert Ernst/Gouws, Rufus H./Wolski, Werner (Hg.): Lexikographica. Berlin/New York, pp. 79-93.

Oi Yee Kwong. 1998. *Aligning wordnet with additional lexical resources*. Proceedings of the COLING-ACL'98 Workshop on 'Usage of WordNet in Natural Language Processing Systems', Montreal, QC, Canada, pp. 73-79.

Kenneth C. Litkowski. 1999. *Towards a meaning-full comparison of lexical resources*. Proceedings of the ACL Special Interest Group on the Lexicon Workshop on Standardizing Lexical Resources, College Park, MD, USA, pp. 30-37.

Christian M. Meyer and Irina Gurevych. 2011. *What psycholinguists know about chemistry: Aligning Wiktionary and WordNet for increased domain coverage*. Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP 2011), pages 883–892,

Elisabeth Niemann and Iryna Gurevych. 2011. *The People's Web meets Linguistic Knowledge: Automatic Sense Alignment of Wikipedia and WordNet*. Proceedings of the Ninth International Conference on Computational Semantics, pp. 205-214.

Roberto Navigli. 2006. *Meaningful clustering of senses helps boost word sense disambiguation performance*. Proceedings of COLING 2006 and ACL 2006. Association for Computational Linguistics, pp. 105-112.

Simone P. Ponzetto and Roberto Navigli. 2010. *Knowledge-rich Word Sense Disambiguation rivaling supervised system*. Proceedings of the 48th Annual Meeting of the ACL, pp. 1522-1531.

Maria Ruiz-Casado, Enrique Alfonseca, and Pablo Castells. 2005. *Automatic Assignment of Wikipedia Encyclopedic Entries to WordNet Synsets*. Advances in Web Intelligence, Volume 3528 of LNCS, Springer Verlag, pp. 380-386.