Helping Swedish words come to their senses: word-sense disambiguation based on sense associations from the SALDO lexicon

Ildikó Pilán
Språkbanken, Dept. of Swedish, University of Gothenburg
Gothenburg, Sweden
ildiko.pilan@svenska.gu.se

Abstract

This paper describes a knowledge-based approach to word-sense disambiguation using a lexical-semantic resource, SALDO. This hierarchically organized lexicon defining senses in terms of other related senses has not been previously explored for this purpose. The proposed method is based on maximizing the overlap between associated word senses of nouns and verbs co-occuring within a sentence. The results of a small-scale experiment using this method are also reported. Overall, the approach proved more efficient for nouns, since not only was the accuracy score higher for this category (56%) than for verbs (46%), but for nouns in 22% more of the cases was a sense overlap found. As a result of an in-depth analysis of the predictions, we identified a number of ways the system could be modified or extended for an improved performance.

1 Introduction

Word-sense disambiguation (WSD) aims at computationally determining the correct sense of a word in a given context (Agirre and Edmonds, 2007). Research in the area began already around the 1950s, being that the successful completion of this task is a prerequisite for a large number of complex Natural Language Processing (NLP) applications (Navigli, 2009). Navigli (2009) as well as Agirre and Edmonds (2007) offer a detailed overview of WSD methods. Such techniques can be classified, among others, based on the amount of knowledge-sources required, i.e. knowledge-based vs. statistical approaches. A wide-spread knowledge-rich method for WSD is the *Lesk algorithm* (Lesk, 1986), based on the overlap between information available about a sense in a lexical resource and the context which the target word appears in. The Lesk algorithm has received a lot of interest and different modified versions of it have also been tested, e.g. Banerjee and Pedersen (2002), Miller et al. (2012), Ekedahl and Golub (2004).

Numerous previous studies deal with WSD for English (Navigli, 2009), but there are significantly fewer examples of WSD for other languages in the literature. The SENSEVAL-2 Workshop (Edmonds and Cotton, 2001) aimed at the extension of WSD to a number of different languages, including a Lexical Sample task for Swedish. Kokkinakis et al. (2001) describe the Swedish data and report the results of the participating systems using this material. The Swedish dataset included altogether 10241 instances selected from the Stockholm-Umeå Corpus (SUC) (Ejerhed and Källgren, 1992) with an average polysemy of 3,5 senses per lexeme. The best performing system for Swedish (Yarowsky et al., 2001) achieved an overall precision of 74,7% for mixed-grained distinctions, where verbs were significantly harder to disambiguate than nouns (a precision of 63,4% compared to 76,9% respectively). A subsequent attempt at Swedish WSD, based on Random Indexing and word co-occurrence, is described in Hassel (2005). This system obtained an accuracy of about 80% on a small dataset comprised of 133 instances, aiming at distinguishing three senses of the word *resa*, namely the senses “journey”, “to travel” and “to raise”.

In the following, we describe the first results obtained with a knowledge-based WSD system under development using *SALDO* (Borin et al., 2013), a Swedish lexical-semantic resource. Exploring this lexicon for WSD is particularly interesting, since its structure differs considerably from WordNet (Fellbaum, 1998), a common alternative employed for this task. Unlike WordNet, SALDO covers all parts of speech (POS) and it is based on association relations among hierarchically organized word senses. Our WSD method builds on the idea that by taking into consideration the overlap between a list of associated senses of nouns and verbs occurring within a sentence, we can disambiguate their senses.

In this paper, we first present SALDO and our test data in section 2. Section 3 provides details about our knowledge-based WSD method, whilst results are presented in section 4. Section 5 includes a thorough analysis of errors and finally, section 6 concludes the paper and outlines directions for further research.

2 Resources used

Our main resource, SALDO was used both as basis for the sense inventory and as source of information about associations between senses. This lexicon contains hierarchically organized word senses where the top node, *PRIM*, is an artificial node whose children consist of 43 core senses. Instead of textual definitions, each sense is defined in terms of another manually selected sense, a mandatory primary descriptor (PD), and one (or more) optional secondary sense descriptors. Each descriptor is a more central semantic neighbor of a given entry. Centrality is determined in terms of frequency, stylistic unmarkedness, morphological complexity and directional semantic relations (e.g. hyperonyms as descriptors of their hyponyms). Due to the structure of SALDO, each sense can be characterized by a a list of *ancestors* (or *semantic path*) consisting of all the primary descriptors encountered while moving upwards in the hierarchy until reaching the top node (*PRIM*). We indicate different SALDO senses with a superscript digit following the word throughout this paper.

In absence of a dataset annotated with SALDO senses for a variety of parts of speech, we evaluated our method on a set of sentences collected via a corpus query tool, *Korp* (Borin et al., 2012b), checking manually whether the predicted senses were correct. Our system made sense predictions for each noun and verb in the sentence, but we only inspected the sense of one target item per sentence. As targets for WSD we used 5 polysemous nouns and equally many verbs suggested by native speakers based on their intuitions. We considered 10 sentences for each item which resulted in 100 test sentences in total. The amount of our test data was limited due to time constraints and the cost of manual sense annotation and error analysis. Our target items and the English equivalent of their first sense (*Eng*) are listed in Tables 1 and 2. The tables include also the number of senses (*# senses*) in SALDO and the average number of senses per POS category (*Avg*).

POS	Lemma	Eng	# senses	Avg
noun	mål	“goal”	7	**4**
	val	“choice”	4	
	glas	“glass”	3	
	ask	“ash tree”	2	
	lov	“permit”	4	

Table 1: Target nouns to disambiguate.

POS	Lemma	Eng	# senses	Avg
verb	läsa	“read”	2	**3.2**
	flyga	“fly”	2	
	ersätta	“substitute”	2	
	spela	“play”	8	
	väcka	“arouse”	2	

Table 2: Target verbs to disambiguate.

The sentences constituting our test set were randomly selected via Korp from the *LäsBarT* corpus (Heimann Mühlenbock, 2013), a collection of easy-to-read newspaper and fiction texts. Since the semantic paths of all nouns and verbs in a sentence were considered by our WSD method when looking for overlaps (without the introduction of a more limited window size) we opted for using a corpus that typically contains shorter sentences than other corpora do, to increase the feasibility of a manual error analysis. Since sentences were randomly selected, the distribution of senses was uneven in the 10 example sentences per lemma.

3 Method

The knowledge-based WSD method proposed relies on maximizing the overlap between the ancestor senses of nouns and verbs appearing within a sentence-long context. By looking at shared senses higher in the SALDO hierarchy, we aim at capturing the idea of semantic relatedness, potentially a shared domain.

In the first step of our WSD, the list of ancestors for each noun and verb in the sentence is collected via *Karp* (Borin et al., 2012a), an online infrastructure for Swedish lexical resources. We access Karp through its web-services using a base form search with a lemma and a POS, provided for each token via the Korp annotation pipeline. Then, the ancestor-overlap for combinations of sense pairs for each pair of lemmas is computed. The comparison does not take place only within the same POS category, i.e. noun and verb senses are compared also to each other. Since for longer paths the absolute number of overlaps would be higher, we introduced a normalization step: the number of overlapping ancestors in the paths of the two senses compared is divided by the summed length of these paths. In a subsequent step, the scores from the pair-wise comparison are summed for each sense and the sense maximizing the overlap with other senses is suggested as disambiguated sense. If no overlap is found, the fallback strategy is choosing the first sense from SALDO. In the case of multi-word expressions (MWE), the corresponding lemma and sense is comprised of more than one word, e.g. *spela_roll* "matter". For such lemmas, the following word is checked in the sentence, in attempt to identify the MWE. If a match is found, the multi-word sense is chosen as prediction.[1]

4 Results

The results of our experiment are presented in Table 3 in terms of the number of correct predictions and fallback predictions for the 10 test sentences per each target item. The amount of correct fallback predictions is indicated in parenthesis, a missing value meaning zero. We also included a baseline accuracy, the average accuracy for nouns and verbs in general (*Avg acc*) and the overall accuracy of the system. We used as baseline our fallback, that is always opting for sense number 1.

The average accuracy of the system over all 100 sentences tested was 51% for disambiguating lemmas with an average polysemy of 3.6, which was only a 1% improvement over the baseline. Although the verbs tested had, on average, almost one sense less to choose from (Table 2) and a higher baseline, the accuracy of our system was 10% lower for verbs than for nouns. In the case of nouns, the system achieved an accuracy of 56% which was 10% above the baseline, whilst for verbs the performance remained 8% below the baseline. Moreover, overlaps were much more common for nouns, where in only 8% of the cases was the fallback of choosing sense number 1 used (in absence of an overlap), out of which none were correct guesses. Verbs tended to create fewer overlaps, 15 out of 50 predictions were fallbacks. For verbs, 20% of correct predictions were obtained with the fallback strategy, which suggests that the overlap-based method proposed is more suitable for nouns. This, however, would need to be confirmed by further experiments on a larger dataset.

Our system's performance compared to the best system in the SENSEVAL-2 task (Yarowsky et al., 2001) remains rather low (17,4% lower for verbs and 20,9% lower for nouns, see section 1). However, the knowledge resource, the sense inventory, as well as the target lemmas and the corpus used were different, which makes a direct comparison hard.

5 Error analysis

To acquire a better understanding about why the system failed to disambiguate senses in certain cases and how the results could be improved, we performed a detailed error analysis of our test sentences. This showed that there are a number of different reasons behind the inaccurate predictions, some of the most common causes being: the insufficient amount of context, the limited number of ancestors, disregarded paradigmatic information, a lack of evidence of common domain or topic, undetected multi-word expressions and the absence of frequency information for a sense.

In the following, we provide an example for some of these categories. Besides the target lemma to disambiguate, we consider also the senses of other nouns and verbs in the sentence that were involved in producing the overlap, highlighting some factors that aided or inhibited successful disambiguation.

One of the potential pitfalls of our approach was the lack of a sufficient amount of context for producing a useful overlap. In the sentence *Sen kommer han på att han behöver en* ***ask***. "Then he remembers that he needs a **box**.", the noun *ask* could be either "box" or "ash tree". Since both verbs are rather generic, none of them produces an overlap with any of the senses of *ask*. In such cases frequency and word co-occurrence informa-

[1]MWE senses were excluded from the counts in Tables 1 and 2 since currently the system cannot detect discontinuous MWE, which results in a rather low MWE recall.

POS	Nouns					Verbs				
Lemma	mål	val	glas	ask	lov	läsa	flyga	ersätta	spela	väcka
# correct predictions	6	5	5	5	7	5	4	4	3	7
# fallbacks (# correct)	1	1	2	0	0	1	2	4 (4)	0	8 (6)
Avg acc (baseline)	56% (46%)					46% (54%)				
Overall acc (baseline)	51% (50%)									

Table 3: WSD results on 100 test sentences.

tion might improve WSD.

Information about the paradigm, i.e. which inflectional pattern a word follows, could also reduce ambiguity in certain cases. In *Nästan hälften av socialdemokraterna i* ***valet*** *tjänar mer än 400 tusen kronor om året.* "Almost half of the socialdemocrats in the **election** earns more than 400 thousand Swedish crowns per year.", the guessed sense for the word *val* was *val*^{2} "whale" (PD: *djur*^{1} "animal"). The base form of this sense (*val*) is the same as that of the correct sense *val*^{4} "choice" (PD: *välja*^{1} "choose"), however, *val*^{2} is a common gender noun, whilst *val*^{4} is of neuter gender. Consequently, the inflected word form in the sentence above, containing the neuter definite ending *-et*, would have been able to rule out *val*^{2}.

There are cases in which the prediction is wrong since information from SALDO is not sufficient for disambiguation, even though the context would be enough for a human to identify a common domain or topic and thus, the correct sense. Consider the example of *Smutsiga grytor, tallrikar och* ***glas*** *trängdes på diskbänken.* "Dirty pots, plates and **glasses** crowded the sink.". Our system used the fallback strategy and guessed *glas*^{1} "glass" (PD: *material*^{1} "material"), instead of the correct solution *glas*^{2} "glass" (PD: *dricka*^{1} "drink") since no overlap was found among the correct senses *gryta*^{1} "pot" (PD: *kärl*^{1} "vessel"), *glas*^{2} "glass" (PD: *dricka*^{1} "to drink") and *tallrik*^{1} "plate" (PD: *mat*^{1} "food"). All these nouns belong to the same topic that could be labeled as *kitchen*, but this is not always reflected in their ancestors. This example would suggest that our system could benefit from the integration of additional information about the domain to which each word sense belongs.

Furthermore, we can find examples where an incorrect prediction could be avoided if a more sophisticated method for detecting multi-word units was used in a step preceding WSD. In *Tiden* ***spelar*** *egentligen inte så stor roll.* "The time does not really **matter** so much.", the verb *spela* and the noun *roll* together form a MWE meaning "matter". SALDO contains a corresponding sense (*spela_roll*^{1}), however, since there are intervening words between the two parts of the expression, our system fails to detect this multi-word unit and, thus, the correct sense. Instead, *spela*^{3} "act" (PD: *teater*^{1}) and *roll*^{1} "part" (PD: *spela*^{3}) is chosen based on the overlap of ancestors.

6 Conclusion and future work

We presented a WSD method for Swedish based on overlap counts between word senses from SALDO, a lexicon previously unexplored for this purpose. We achieved 51% accuracy on a dataset with 3.6 average senses per lemma. We found that this approach was more successful for disambiguating nouns than verbs both in terms of accuracy (56% vs 46%) and the amount of overlap found (92% vs 70% of the test items respectively). A detailed error analysis showed that incorporating, among others, strategies handling the absence of overlap and information about topics or domains in this approach could contribute to achieving a more accurate performance. Addressing these areas of improvement could make this WSD system more useful for several NLP tasks including, but not limited to machine translation, sentiment analysis and summarization.

In the future, a number of directions for extending our method could be investigated such as considering secondary descriptors in the overlap counts and taking into consideration dependency relations during disambiguation. The performance of our system would need to be measured on a larger dataset labeled by multiple annotators and disambiguating the sense of adjectives and adverbs with this method could also be explored. Integration with other knowledge resources and vector space models may also be interesting directions to pursue.

References

Eneko Agirre and Philip Glenny Edmonds. 2007. *Word sense disambiguation: Algorithms and applications*, volume 33. Springer.

Satanjeev Banerjee and Ted Pedersen. 2002. An adapted Lesk algorithm for word sense disambiguation using WordNet. In *Computational linguistics and intelligent text processing*, pages 136–145. Springer.

Lars Borin, Markus Forsberg, Leif-Jöran Olsson, and Jonatan Uppström. 2012a. The open lexical infrastructure of Språkbanken. In *LREC*, pages 3598–3602.

Lars Borin, Markus Forsberg, and Johan Roxendal. 2012b. Korp-the corpus infrastructure of Språkbanken. In *LREC*, pages 474–478.

Lars Borin, Markus Forsberg, and Lennart Lönngren. 2013. SALDO: a touch of yin to WordNet's yang. *Language Resources and Evaluation*, 47(4):1191–1211.

Philip Edmonds and Scott Cotton. 2001. SENSEVAL-2: overview. In *The Proceedings of the Second International Workshop on Evaluating Word Sense Disambiguation Systems*, pages 1–5. Association for Computational Linguistics.

Eva Ejerhed and Gunnel Källgren. 1992. The linguistic annotation of the Stockholm-Umeå Corpus project. Technical report, University of Umeå.

Jonas Ekedahl and Koraljka Golub. 2004. Word sense disambiguation using Wordnet and the Lesk algorithm. *Projektarbeten 2004*, page 17.

Christiane Fellbaum. 1998. *WordNet*. Wiley Online Library.

Martin Hassel. 2005. Word sense disambiguation using co-occurrence statistics on random labels. In *Proceedings of Recent Advances in Natural Language Processing 2005*, Borovets, Bulgaria.

Katarina Heimann Mühlenbock. 2013. *I see what you mean*. Ph.D. thesis, University of Gothenburg.

Dimitrios Kokkinakis, Jerker Järborg, and Yvonne Cederholm. 2001. Senseval-2: the Swedish framework. In *The Proceedings of the Second International Workshop on Evaluating Word Sense Disambiguation Systems*, pages 45–48. Association for Computational Linguistics.

Michael Lesk. 1986. Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In *Proceedings of the 5th annual international conference on Systems documentation*, pages 24–26. ACM.

Tristan Miller, Chris Biemann, Torsten Zesch, and Iryna Gurevych. 2012. Using distributional similarity for lexical expansion in knowledge-based word sense disambiguation. In *COLING*, pages 1781–1796.

Roberto Navigli. 2009. Word sense disambiguation: A survey. *ACM Computing Surveys (CSUR)*, 41(2):10.

David Yarowsky, Silviu Cucerzan, Radu Florian, Charles Schafer, and Richard Wicentowski. 2001. The John Hopkins SENSEVAL-2 system descriptions. In *Proceedings of SENSEVAL-2 Second International Workshop on Evaluating Word Sense Disambiguation Systems*, pages 163–166, Toulouse, France, July. Association for Computational Linguistics.