Semantic Mapping for Lexical Sparseness Reduction in Parsing

Simon Šuster
University of Groningen
Netherlands
s.suster@rug.nl

Gertjan van Noord
University of Groningen
Netherlands
g.j.m.van.noord@rug.nl

Abstract

Bilexical information is known to be helpful in parse disambiguation, but the benefit is limited because of lexical sparseness. An approach using word classes can reduce sparseness and potentially leads to more accurate parsing. Firstly, we describe a method identifying the dependency types of the Alpino parser for Dutch to which we would like to apply generalization. These are the types which are most likely to reduce the sparseness and positively affect parsing at the same time. Secondly, we provide preliminary results for enhancement of dependency types with semantic classes derived from a WordNet-like inventory for Dutch. Classes of varying degrees of generality are applied to three dependency types: nominal conjunction, modification of adjective and modification of noun. We observe improvements in some concrete cases, whereas the overall parsing accuracy either remains unchanged or decreases. We identify drawbacks of human-built sense inventories, which provides motivation for a distributional semantic approach.

1 Introduction

Many modern syntactic parsers use lexicalized information in either strict parsing or parse disambiguation (Bikel, 2002; Collins, 2003; Charniak, 2000; Van Noord, 2006). A bilexical feature, such as the one to learn that a particular verb occurs with a specific object, is modeled from large corpora, yet for many instances during parsing, such bilexical information is missing. To address lexical sparseness, generalization is needed, which can be representative of several lexical units[1] at once and which can lessen the excessive differentiation at the lexical level. We are interested in approaches in which the obtained generalization is semantic. Most work up to date has tried to achieve semantic representation by using sets of senses from external human-built lexical inventories, such as WordNet (Agirre et al., 2008; Henestroza Anguiano and Candito, 2012; MacKinlay et al., 2012), with varying degrees of success.

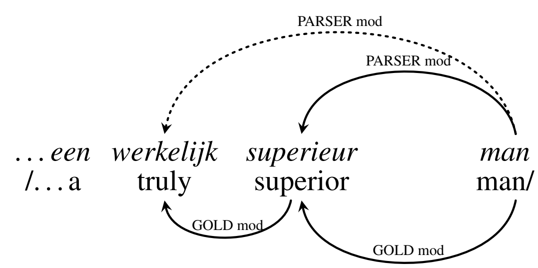

Figure 1: Dependency analyses of the gold annotation (bottom) and the parser (top). The dotted line denotes the wrong attachment.

With this research, we aim at 1) empirically identifying dependency types for which generalization is most needed in parsing; 2) providing baseline results using semantic classes extracted from a human-built inventory; 3) finding out weak aspects of using such inventories, especially in contrast to a distributional approach.

We illustrate the need for generalization and the application of a semantic class on the Dutch sentence excerpt in Figure 1. The parser was unable to attach the adverb *werkelijk* correctly because the

[1] By ’lexical unit’ we mean a form-meaning pair.

pair ⟨*werkelijk*, *superieur*⟩ was missing in the lexical association model. Such cases of modification of adjective could be successfully resolved if we were able to abstract away from concrete words. Indeed, in the Cornetto lexical semantic database for Dutch, the word *superieur* shares the semantic class, or the synset[2], with the word *goed* 'good'. Since the model includes information about ⟨*werkelijk*, *goed*⟩, the semantic enhancement would mean a successful model look-up in the future and possibly a correct parse as well.

Previous studies on this topic have typically focused on enhancing words regardless of the dependency type in which they appear (section 2). Our approach starts by identifying types where the information from the model is particularly scarce and where parsing accuracy is reduced. Following a statistical analysis, we list three types that are likely to be helpful once the semantic information is added to the model (section 4.1).

We experiment with three types of semantic representations (section 3.6) with varying degree of generality. The enhanced model is consulted whenever there is no bilexical preference found in the standard model.

We show that our method leads to improved parses in some cases, but not to an increase in overall parsing accuracy. Further, we show that increased generalization introduces a greater number of parse modifications, but at the expense of precision.

2 Related Work

Research on parsing improvement with generalization can be roughly divided into approaches that introduce lexical semantic information from human-built resources as opposed to those which acquire classes distributionally, e.g. by clustering. Here, we introduce studies from the former type as it is the only one relevant to present work.

One of the earliest attempts in semantically enriched parsing is the work of Bikel (2000), who incorporated WordNet classes into a lexicalized generative PCFG model, but without significant improvement. Bikel considers several levels of generality but does not try to determine a single level.

For Chinese, Xiong et al. (2005) used resources equivalent to WordNet in order to obtain sense information and three levels of generalization (the immediate class together with two hypernym classes). The new information is incorporated as a sub-model in a generative lexicalized model, which improved over their baseline model. Fujita et al. (2007) develop a parse selection model for Japanese, which in its best configuration uses both syntactic and semantic features. The latter are based on dependencies extracted from semantic representations of sentences. Elements in dependency triples are substituted by senses and hypernym classes at various levels. Based on semantic dependencies and valency features, they achieved a substantial improvement over their best syntactic model.

Agirre et al. (2008) experiment with two lexicalized parsing models and map semantic classes to the training data to evaluate on a general parsing task, as well as on PP-attachment disambiguation. Three levels of semantic representation were incorporated: synsets, coarse semantic files and hybrid word–semantic-file representations. The semantic files and the first-sense heuristic for disambiguation turned out to be good performers in most of the experiments. The maximum performance gain was 1.1% (in F-score) in general parsing and 5.6% in PP-attachment. Agirre et al. (2011) reproduced the results by including semantic classes as features into the MaltParser (Nivre, 2006).

Recently, Henestroza Anguiano and Candito (2012) introduced probabilistic features for capturing generalization. Replacement with first-sense synsets from French EuroWordNet yielded slight, though not significant improvements on the French Treebank and an out-of-domain medical corpus. Since the results were better when testing out of domain, they argued that parsing improvement is more likely to be successful when there is a big lexical divide between the training set and the testing set. Similarly, MacKinlay et al. (2012) provide mixed results in HPSG parse selection using the English Resource Grammar. The authors observed no improvement in including synsets from the hypernym path as features. The best performer was the semantic file representation, which reduced the error rate by 1%.

Clark (2001) learnt selectional preferences based on classes obtained from WordNet and evaluated

[2] Synset is a set of lexical units.

them in parse selection. Although he failed to improve the base selection component of the parser of Carroll and Briscoe (1996), the work is pertinent to ours as it addresses the issue of selecting a suitable level of generalization. In contrast to previously cited work, which deals with the problem in an exploratory manner, the procedure involves a chi-square test with significance level acting as a parameter for controlling the extent of generalization.

Compared to previous work, we do not apply semantic classes indiscriminately to words in all dependency types, but enhance only specific dependency types that are identified empirically. We observe that bilexical information is differently useful for different dependency types, which are in turn differently problematic for the parser. Additionally, we believe that this is likely to be parser dependent. This constraint will avoid applying the new information to dependency types for which the bilexical model performs well. Our work further differs by including semantic information not into full parsing, but in the existing parse selection component. More precisely, we enhance only bilexical preferences, which limits the room for improvement.

3 Methodology

In this section, we first present the parser and the resources for training, testing and semantic class extraction. We go on to introduce the method for analyzing the bilexical model usage. We then discuss the disambiguation method, levels of semantic representation, and introduction of the new information in the model and the parsing process.

3.1 Parser

The Alpino parser is a linguistically motivated wide-coverage parser for Dutch, which uses a large lexicon and a large set of HPSG (and other) rules (these are augmented in such a way to represent dependency structure) (Van Noord, 2006). The system employs a MaxEnt model as parse selection criterion, one part of it being the bilexical component ("the model") which verifies the degree of association between a pair of words in a specific dependency relation. It was shown in Van Noord (2010) that the incorporation of this bilexical information into parsing improves over the parser without access to it.

The model describes dependency instances with 35 features, each describing words in different relations and with different parts of speech (POS). For the example from Figure 1, one of the applicable features would be modification of a noun (⟨mod, noun⟩), instantiated as:

(1) ⟨superieur, adj, mod, noun, man⟩
$\langle w_{dep}, \quad pos_{dep}, rel, \quad pos_{head}, w_{head}\rangle$

Every instance in the model is associated with a normalized pointwise mutual information score with a frequency threshold for inclusion of > 50. The score, together with feature's weight, is used when selecting the best parse.

We evaluate parser performance with concept accuracy (CA), which is in practice very similar to labeled attachment score (LAS) (Van Noord, 2006).[3]

3.2 Corpora

The identification of relevant dependency types was carried out on Lassy Small, a hand-annotated corpus of Dutch containing around 1 million words (1.3 million dependency instances) for a large variety of texts (newspapers, Wikipedia, websites, fiction etc.) (Van Noord, 2009). An automatically syntactically annotated part of the preliminary version of Lassy Large, amounting to 500 million words, was used for training the lexical association model of Alpino. We tested our method on the Alpino Treebank (7,136 sentences), which is a collection of newspaper texts from the Eindhoven corpus, and parts of Lassy Small (3,917 sentences).

3.3 Lexical Semantic Database

Cornetto is a lexical semantic database for Dutch and is the result of a merge of Referentie Bestand Nederlands (RBN, a collection of lexical units) and Dutch WordNet (Vossen et al., 2013). It includes more than 92,000 form-POS pairs, described in terms of lexical units, synsets and other criteria. Table 1 lists some statistics of Cornetto, version 2.

3.4 Identification of Dependency Types

An indication of the space for improvement of the model can be obtained from the proportion of test

[3]CA is a mean of per-sentence minimum of recall and precision. The main reason for using this measure is to relax the single-head constraint of LAS.

Type	All	Nouns	Verbs	Adjectives	Adverbs
Synsets	70,370	52,845	9,017	7,689	220
Lexical Units	119,108	85,449	17,314	15,712	475
Form-POS pairs	92,686	70,315	9,051	12,288	1,032

Table 1: Cornetto statistics

instances for which bilexical preference was found in the model (referred to as model coverage). Not all dependency types are equally helpful in parse selection, so we correlate model coverage with parsing performance. On the level of a dependency instance, the accuracy of a parse is treated in a binary fashion. This captures our assumption that dependency types which 1) have a low model coverage and for which 2) failed model look-up negatively affects parsing outcome, are typically promising for semantic enhancement.

3.5 Disambiguation method

To circumvent the problem of sense ambiguity when mapping classes to word forms, we always choose the first listed sense in our database. The first-sense heuristic was shown to be a well performing technique in Agirre et al. (2008; 2011) and McCarthy et al. (2004).

Senses in Cornetto are ranked mostly according to their prominence, reflecting various lexicographic criteria in RBN, one of them being observed frequency in corpora. We point out that the first-sense heuristic applied to Cornetto selects somewhat different information than in the case of English WordNet because the latter is based exclusively on frequency (McCarthy et al., 2004).[4]

For the purpose of determining generality of a synset, we treat Cornetto as a digraph, with nodes constituting synsets and arcs constituting hypernymic relations. Since the graph resembles a tree, we use the term *leaves* to denote the most concrete synsets with no incoming arcs (hyponyms), and *top* to denote the most abstract node with no outgoing arcs. We use Information Content (IC) as defined in (Sánchez et al., 2011) and shown in Equation 1.

$$IC(s) = -log\frac{\frac{|leaves_s|}{|subsumers_s|} + 1}{total\_leaves + 1} \qquad (1)$$

[4]WordNet senses not observed in the gold are ordered arbitrarily.

IC of a synset s is a function of the cardinality of s's leaves, the cardinality of s's subsumers and the total number of leaves in Cornetto.

3.6 Semantic representation

We decide to use immediate synset of a form-POS pair as a fine-grained semantic representation of a lexical unit. The application of word senses in parsing actually introduces a duality: on the one hand, the sense information is more specific than a form-POS pair (which can have one or more senses); on the other hand, synsets consist of lexical units, thus providing an abstraction (smallest possible) from a lexical unit (cf. Bikel (2000)).

A much coarser representation is possible through semantic types, which provide a very general description analogous to English WordNet semantic files, or super-senses (MacKinlay et al., 2012; Agirre et al., 2008). There are around 20 types in Cornetto, assigned to approximately half of lexical units. The eight most frequently assigned types are:

(2) nondynamic, action, artefact, human, dynamic, abstract, place, concrete

These are POS-dependent, so, for example, a verb could be assigned the "action" class, but not "place", which is reserved for nouns and adjectives.

An intermediate degree of granularity is achieved by using Information Content introduced in section 3.5. For a given immediate synset of a lexical unit, s, we look up its IC value. If $IC(s)$ exceeds the threshold δ, s is too concrete, and a more general synset needs to be considered. The suitable generalized synset s_{gen} is the one closest to, but below δ. Conversely, when $IC(s) < \delta$, the synset is already located high in the hierarchy, therefore no generalization is needed. We set the value of δ manually, by inspecting hypernymic paths of various lexical units. At this stage of research, no empirical optimization of δ is attempted. For the experiment described in the following, δ was set to 6.

3.7 Mapping to Model and Parsing Integration

Incorporating semantic classes into the model is straightforward. Firstly, we retrieve instances of identified dependency types from the bilexical model, then we map class identifiers to words in these instances. Finally, the MI scores are calculated. In order to enable mapping between word forms in the model and information from Cornetto, the Cornetto forms are first analyzed with the Alpino lemmatizer.

The new information is made available as additional data to the standard model. This relates to the way how the parser uses semantic classes. Semantically enhanced instances are only considered after a failed look-up in the bilexical model.[5] The reason for using a back-off strategy relates to how we perceive our task. Bilexical preferences improve parsing as shown by Van Noord (2010). Here, we see generalization as potentially helpful where the problem of lexical sparseness is too severe. Intuitively, lexical information is more precise than generalized-class representation, so the former should be used whenever possible to keep the precision high.

4 Results

4.1 Relevant Dependency Types

Table 2 displays lexical model coverage and variability in proportion of instances for which the model was able to provide bilexical information. The coverage is rarely very high, except for cases such as verbal complements, ⟨verb, vc⟩, where the complement can only be introduced with a limited set of words.

We correlate the model coverage with parsing accuracy by simply taking into account whether the parser was correct on a particular instance or not. The Pearson's chi-square test confirms ($p < 0.001$) that parsing accuracy on instances which *were* in the model differs to instances *not* in the model. On all dependency types, an incorrect parse is 3.65 times (odds) more likely when the instance is not in the model.

The dependency types for improvement are selected according to the following three criteria: odds ratio, Cramer's ϕ correlation coefficient (both measuring effect size) and number of out-of-model instances that were parsed incorrectly. Note that all types have relatively low correlation coefficients because the bilexical information is only one out of many feature types used in parse selection. For our purpose, the prevailing error type observed in incorrect parses should be wrong attachment. We therefore manually discard dependency types which look promising based on high scores for the listed criteria, but in reality mostly include other error types, such as incorrect relation labels resulting from limitations in the grammar, or non-standard language phenomena. The final selection consists of modification of the adjective, nominal coordination and modification of the noun (see Table 3). We map synsets to nouns, adjectives and adverbs occurring in these identified types, and semantic types to nouns and adjectives as they are not defined for adverbs.

Type	# Lassy Small	# Model	%
(verb, vc)	41798	40877	97.8
(verb, mod)	96609	87591	90.7
(prep, obj1)	135645	115582	85.2
(verb, su)	113658	86640	76.2
(noun, mod)	133925	86430	64.5
(noun, cnj)	18848	5512	29.2
pp(adj, mod)	3254	569	17.5
all	924783	672535	72.7

Table 2: Model coverage for a selection of dependency types

Type	# instances	Odds	ϕ coef.	Out-of-model misparsed instances #	Out-of-model misparsed instances %
(adj,mod)	11828	2.653	0.2	1213	10.3
(noun,cnj)	18853	2.042	0.12	3192	16.9
(noun,mod)	133925	1.962	0.11	8003	6.0

Table 3: Identified dependency types with respective statistics

4.2 Semantic Classes

The configuration using immediate synsets (SYN) does not result in an overall improvement of parsing accuracy. The average performance with ten-fold cross validation levels the baseline parser configuration at 90.46% CA. Out of the cases which were mapped to Cornetto successfully and which were also found in the enhanced model, SYN leads to 33 improved dependency instances, as opposed to 29 instances where the accuracy deteriorated (Table 4). There are several reasons why the number of actual

[5] We also experimented with using semantic classes by default and backing off to bilexical model, but this led to reduced performance.

parse modifications here is small. Firstly, the access to the new information is attempted relatively rarely, only after a failed model look-up. Secondly, Cornetto synsets could only be mapped to around 60% of the form-POS pairs encountered during parsing. Thirdly, the success of finding an enhanced instance having at least one successfully mapped synset ID is 7.85%. This low number means that synsets cannot generalize sufficiently, which is understandable – many synsets include only one lexical unit.

In the experiment with coarse semantic types (ST), the results are worse, and the parsing accuracy drops to 90.35%. Although the number of correctly introduced parses increases in comparison with SYN (to 178), the number of modifications which result in an inaccurate parse increases as well (299). The new information thus clearly overgeneralizes. It is possible that the generalization we obtain with only 20 semantic types for Dutch, as opposed to 45 types in English WordNet (Agirre et al., 2008), influences the resulting precision. The results in Table 4 show how a better recall (*% found*) from the enhanced model corresponds to a deteriorated precision (*# deteriorated*).

The best performing type among the 3 identified types was nominal conjunction. We applied the method of intermediate granularity (INT) to this type only. Results for nominal conjunction are shown in Table 5. INT method does introduce a higher number of modifications than SYN, but again at the expense of precision. Further experiments would be needed in order to confirm whether INT is a better performer than the ST method.

A manual inspection of instances that were parsed correctly thanks to the incorporated semantic classes confirms our reasoning around the motivating example from the introduction (Figure 1) for all three dependency types. Consider the following sentence encountered during testing:

(3) De afgelopen week werden er in de **Utrechtse Camera bioscoop** elke dag Tarzanfilms gedraaid.
Last week, Tarzan films were shown every day in the Camera cinema in Utrecht.

The attachment of the word *Utrechtse* is ambiguous between *Camera* and *bioscoop*. The pair representing the correct attachment, $\langle$*Utrechtse*, *bioscoop*$\rangle$, was not found in the bilexical model and was not attached correctly by the parser. However, an instance in which *Utrechtse* was substituted with a synset representing place was present in the enhanced model, which enabled a successful parse. In contrast, incorrect parses are mostly introduced either because the head–dependent pair for which the new model provides support stands for a wrong attachment, or because the relation label in gold is different (e.g. apposition instead of modification).

		#		CA	
method	% found	improved	deteriorated	new	old
SYN	7.8	33	29	90.46	90.46
ST	62.1	178	299	90.35	90.46

Table 4: Enhancement results on the 3 identified dependency types (see Table 3). *SYN*: immediate synset, *ST*: semantic type, *found*: instances found in the enhanced model when backing off, *improved*: incorrect-to-correct parse modification, *deteriorated*: correct-to-incorrect parse modification. Total number of test sentences: 11,053.

	#		CA	
method	improved	deteriorated	new	old
SYN	7	2	90.47	90.46
ST	20	26	90.45	90.46
INT	16	19	90.45	90.46

Table 5: Enhancement results on $\langle$noun, cnj$\rangle$ type only. *INT*: class of intermediate granularity (based on IC scores with $\delta = 6$).

5 Conclusion

We have presented preliminary results for measuring and reducing lexical sparseness with semantic classes, applied to parsing of Dutch. The enhancement of the bilexical model for parse selection with semantic classes did not result in improved overall parsing accuracy, which somewhat mirrors the results of MacKinlay et al. (2012) and Bikel (2000). In the experiment in which immediate synsets are used as back-off after a failed lexical look-up, the number of correct parses is higher than the number of incorrect ones, but the total number of modifications is low.

Because they are fine-grained, synsets do not effectively reduce sparseness. In the light of previous research, it is surprising that the coarse semantic types performed worse than synsets. This could be at least partly attributed to a different specification and number of semantic types in English WordNet and Dutch Cornetto. The experiment in which we choose the suitable level of representation by measuring synset generality gave slightly better results than that of semantic types, although further work should be performed to assess the effect of the threshold parameter on the results.

The fact that the number of parse modifications is low can be explained by Cornetto coverage and the back-off strategy. Our results suggest that higher levels of generalization yield a higher number of parse deteriorations.

One of the factors which might contribute to the rather disappointing results could be selection of incorrect senses for a given word form. It is possible that the first most prominent sense in Cornetto is often not the most frequent sense. Currently, the form of Dutch SemCor which would allow us to reorder Cornetto senses by frequency does not yet exist (Vossen et al., 2013). This is unlikely to matter for senses which are difficult to distinguish and for high levels of representation with higher chance of converging senses. Further, as noted at the end of section 2, the room for improvement in our experiments is smaller compared to studies enhancing full parsing or selection components more generally. It seems that the degree of lexicalization of the parser partly determines the impact of generalization techniques too. For example, Plank (2011) shows that removing lexical features from Alpino's selection component affects the performance relatively little compared to some data-driven parsers whose performance can drop as much as 10.5% when unlexicalized. The drop percentage might be an indicator of the expected final impact of generalization.

Next, we plan to develop a distributional model for inducing semantic classes, which could effectively tackle the problem of resource coverage – note that around 40% of words could not be mapped to semantic classes because they were not found in the lexical semantic inventory. Distributional semantic approaches have further advantages, namely increased adaptability for parsing out of domain, possibility to vary sense granularity and sense ranking, and sense disambiguation when composing meaning representations.

References

Eneko Agirre, Timothy Baldwin, and David Martínez. 2008. Improving Parsing and PP Attachment Performance with Sense Information. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics*, ACL 2008, pages 317–325.

Eneko Agirre, Kepa Bengoetxea, Koldo Gojenola, and Joakim Nivre. 2011. Improving dependency parsing with semantic classes. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, HLT '11, pages 699–703.

Daniel M. Bikel. 2000. A statistical model for parsing and word-sense disambiguation. In *Proceedings of the Joint SIGDAT conference on Empirical methods in natural language processing and very large corpora*, EMNLP '00, pages 155–163.

Daniel M. Bikel. 2002. Design of a multi-lingual, parallel-processing statistical parsing engine. In *Proceedings of the second international conference on Human Language Technology Research*, HLT '02, pages 178–182.

John Carroll and Ted Briscoe. 1996. Apportioning development effort in a probabilistic lr parsing system through evaluation. In *Proceedings of the Empirical Methods in Natural Language Processing, EMNLP-96*, pages 92–100.

Eugene Charniak. 2000. A maximum-entropy-inspired parser. In *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference*, NAACL 2000, pages 132–139.

Stephen Clark. 2001. *Class-Based Statistical Models for Lexical Knowledge Acquisition*. Ph.D. thesis, University of Sussex.

Michael Collins. 2003. Head-driven statistical models for natural language parsing. *Computational Linguistics*, 29(4):589–637.

Sanae Fujita, Francis Bond, Stephan Oepen, and Takaaki Tanaka. 2007. Exploiting semantic information for HPSG parse selection. In *Proceedings of the Workshop on Deep Linguistic Processing*, DeepLP '07, pages 25–32.

Enrique Henestroza Anguiano and Marie Candito. 2012. Probabilistic lexical generalization for French dependency parsing. In *Joint Workshop on Statistical Parsing and Semantic Processing of Morphologically Rich Languages*, SP-Sem-MRL 2012, pages 1–11.

Andrew MacKinlay, Rebecca Dridan, Diana McCarthy, and Timothy Baldwin. 2012. The effects of semantic annotations on precision parse ranking. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics*, SemEval '12, pages 228–236.

Diana McCarthy, Rob Koeling, Julie Weeds, and John A. Carroll. 2004. Finding predominant word senses in untagged text. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, pages 279–286.

Joakim Nivre. 2006. *Inductive Dependency Parsing (Text, Speech and Language Technology)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Gertjan van Noord. 2006. At Last Parsing Is Now Operational. In *TALN06. Verbum Ex Machina. Actes de la 13e conference sur le traitement automatique des langues naturelles*, pages 20–42.

Gertjan van Noord. 2009. Huge parsed corpora in LASSY. In *Proceedings of TLT7*, LOT, pages 115–126.

Gertjan van Noord. 2010. Self-trained bilexical preferences to improve disambiguation accuracy. In *Trends in Parsing Technology: Dependency Parsing, Domain Adaptation, and Deep Parsing*, Text, Speech and Language Technology, pages 183–200.

Barbara Plank. 2011. *Domain Adaptation for Parsing*. Ph.D. thesis, University of Groningen.

David Sánchez, Montserrat Batet, and David Isern. 2011. Ontology-based information content computation. *Knowledge-Based Systems*, 24(2):297–303, March.

Piek Vossen, Isa Maks, Roxanne Segers, Hennie van der Vliet, Marie-Francine Moens, Katja Hofmann, Erik Tjong Kim Sang, and Maarten de Rijke, editors, 2013. *Cornetto: A Combinatorial Lexical Semantic Database for Dutch*. Springer.

Deyi Xiong, Shuanglong Li, Qun Liu, Shouxun Lin, and Yueliang Qian. 2005. Parsing the Penn Chinese treebank with semantic knowledge. In *Proceedings of the Second international joint conference on Natural Language Processing*, IJCNLP'05, pages 70–81.