Using Semantic Features Derived from Word-Space Models for Swedish Coreference Resolution

Kristina Nilsson and Hans Hjelm
Swedish Graduate Schoool of Language Technology (GSLT) and
Computational Linguistics, Department of Linguistics
Stockholm University, SE-106 91 Stockholm
{kristina.nilsson,hans.hjelm}@ling.su.se

Abstract

We investigate the effect of using word-space models as an approximation of the kind of lexico-semantic and common-sense knowledge needed for coreference resolution of definite descriptions, that is, definite NPs with a common noun as head, for Swedish news text. We contrast a system using semantic knowledge from the word-space models with a semantically ignorant system and another system drawing its semantic information from a semantic dictionary called SynLex. We demonstrate an improvement in the results for two different evaluation tasks for the system using word space-derived semantic information over both other systems.

1 Introduction

Coreference resolution, that is, the identification of all expressions referring to the same entity within a discourse, is an important preprocessing step in many Natural Language Processing tasks, for example question answering, information extraction, automatic summarization, and machine translation (Mitkov, 2003). For example, extrinsic evaluations of the effect of adding coreference resolution to systems for question answering show that adding referential relationships between noun phrases improves system performance as well as the quality of retrieved answers for passage retrieval (Morton, 2005), and that the coverage of off-line answer extraction is improved (Hendrickx et al., 2008a).

The coreference resolution task, when applied to noun phrases, can be further divided into the following sub-tasks where the classification is based on the type of referring expression:

a) pronoun resolution, e.g., the pronoun 'he' can be used to refer to the NP 'president Kennedy' with the Named Entity (NE) 'Kennedy' as head,

b) identification of coreferent NEs, e.g., 'John F. Kennedy', 'Kennedy', 'President Kennedy', and 'JFK' might all refer to the same discourse entity,

c) resolution of definite descriptions, that is, anaphoric definite NPs with a common noun as head, e.g., 'the president of the United States' might refer to the same entity as 'the president' or 'the commander-in-chief' within a discourse.

This paper is concerned with the task 'c', the resolution of coreferent definite descriptions. This is a challenging problem in comparison to Named Entity coreference resolution ('b') and pronoun resolution ('a'). For example, (Strube et al., 2002) report an f-score of 33.94% for definite description resolution using a knowledge-poor, language- and domain-independent approach. The results for definite descriptions are markedly lower than the results for NEs and pronouns (with f-scores of 76.22% and 81.60% respectively) as well as the overall result for the system (an f-score of 67.89%).

But however difficult, it is an important task: in the coreference annotated data used in this experiment, 24% of all subsequent-mention coreferent NPs are pronouns, 32% are NEs, and 44% are definite descriptions. Further, resolution of definite descriptions might be of interest in information access tasks such as information extraction and question answering because definite descriptions carry additional information about the discourse entity in question, for example that the entity denoted by the NE 'John F. Kennedy' in some discourse also is referred to by the definite description 'the president of the United States'.

Kristiina Jokinen and Eckhard Bick (Eds.)
NODALIDA 2009 Conference Proceedings, pp. 134–141

Resolution of definite descriptions in turn includes a number of sub-tasks of varying difficulty; we distinguish between these tasks:

1) *resolution of identical head definite descriptions*: cases where the anaphoric definite description and the antecedent share the same head noun, as in the following example: 'She has a **revenue** of three million a year [...]. The **revenue** of Elly Lagerin's store ...'[1]
About 50% of all anaphoric definite descriptions in our data share the same head noun as the antecedent, and thus can be resolved with various string and substring matching techniques combined with morphological analysis;

2) *resolution of non-identical head definite descriptions*: the remaining 50% of all anaphoric definite descriptions are cases where the anaphor has a different head noun than the antecedent. We distinguish between two types of cases based on whether the head of the antecedent NP is a NE or a common noun:

    a) In cases where the antecedent is a NE of a certain type and the head noun of the anaphor is a common noun, as in the antecedent-anaphor pair '<NE type='PERSON'>Hans Stråberg</NE>' - 'the CEO of Electrolux', an estimate of the semantic compatibility of the candidate antecedent and the anaphor might help resolution,

    b) In cases where both the anaphor and the antecedent are definite descriptions but their head nouns are non-identical, resolution might depend on information on lexical relations such as synonymy, hypernymy or hyponymy, or on additional information required for further reasoning and/or keeping track of the current focus.

The main topic of this paper is resolution of non-identical head definite descriptions. We describe an experiment on modeling lexical knowledge on domain-specific data using word-space models. This knowledge is used for deriving features for coreference resolution of candidate antecedent-anaphor pairs. In order to evaluate these semantic features, they are added to a baseline feature set consisting of morphological, lexical, positional, and syntactic features. We also compare the effect of the word-space features to the effect of features based on a semantic dictionary, SynLex.

While coreference resolution is an important preprocessing task for many NLP tasks, the availability of resources needed for the task varies depending on the language and the domain. For the sub-task of resolution of definite NPs with a common noun as head, information on semantic relatedness is essential. The word-space model meet these needs well: it can provide lexico-semantic similarity judgements in any language and domain, as long as the appropriate text material is available. This is our main reason for choosing to work with word-space, or *distributional*, semantics in our experiments.

2 Related Work

Systems for coreference resolution (either for the coreference problem as a whole, or focusing on sub-tasks such as pronoun resolution, or processing of anaphoric definite NPs with common noun heads) commonly use resources such as the lexical database WordNet (Fellbaum, 1998) or its (smaller) European counterparts in EuroWordNet (Vossen, 1998) for adding information on semantic relatedness between NPs.

For example, WordNet was used to test the semantic compatibility of individual NP pairs by assigning the first WordNet sense of the head noun as the semantic class of common noun NPs by (Soon et al., 2001), who found that both a better algorithm for assigning semantic classes and a more refined semantic class hierarchy were needed.

(Ng, 2007) shows that a system for English using automatically induced semantic class knowledge performs better than a system using the WordNet first sense heuristic, while (Hendrickx et al., 2008b) reports that combining features based on automatically generated semantic clusters with features based on synonym and hypernym relations in Dutch EuroWordNet, gives a small but significant improvement.

Other studies have also shown that the knowledge encoded in WordNet is insufficient for coreference resolution, e.g., there are limitations as to

[1]This example is an approximate translation from our Swedish data.

coverage of both vocabulary and relations, ambiguity (there might be more than one sense to a concept, and synsets in WordNet are sorted by frequency), and semantically related words might be located far from each other in the WordNet structure (see e.g., (Vieira and Poesio, 2000; Poesio et al., 1998)).

Furthermore, WordNet is a general ontology, while resolution might require domain-specific or context-dependent lexical information. Efforts towards automatically acquiring such information from corpora are described by e.g., (Poesio et al., 1998; Goecke et al., 2007). Again, as mentioned in Sect. 1, we choose to work with word-space semantics, precisely for its ability to provide language and domain-specific lexico-semantic knowledge to our system.

3 Semantic Features for Coreference Resolution

In this experiment, coreference is defined as a relation of identity of reference between two noun phrases. The resolution task is limited to classification of pairs of possibly anaphoric NPs and their candidate antecedents; the subsequent linking of classified pairs into coreference chains will not be discussed here as the aim of the paper is to discuss the influence of semantic features on the classification task.[2]

The task is further limited to resolution of non-identical head anaphora (listed as type '2' in Sect. 1), i.e., cases where we cannot rely on string matching for resolving the anaphoric reference. We also divide the pair-wise classification into two sub-tasks, based on the respective NP types of the candidate antecedent and anaphor:

1. the candidate anaphor is a definite NP with a common noun head, and the candidate antecedent is a NE – listed as '2a' in Sect. 1;
2. the candidate anaphor and the candidate antecedent are both definite NPs with non-identical common nouns as head – listed as '2b' in Sect. 1.

3.1 Semantic relatedness as expressed in SynLex

SynLex[3] is a free dictionary of general vocabulary Swedish synonyms consisting of 25.000 word pairs (Kann and Rosell, 2006). Synlex was automatically constructed and later manually refined by volunteer users of an on-line dictionary. The users graded each candidate synonym pair according to their intuitive estimate as to how closely the candidate pair was related (semantically), and pairs with a user grade above a certain threshold were included in the dictionary. For each pair of words in SynLex, there is a score between 3.0 and 5.0 representing how the users graded the pair. According to (Kann and Rosell, 2006), pairs with a score of 3.0 are synonymic to a lesser degree, whereas pairs with a score of 4.0 are very good synonyms. SynLex, unlike WordNet, does not distinguish between different word senses.

We use SynLex for deriving two relational features, one binary feature indicating whether the base form of the head word of the candidate antecedent and the base form of the head word of the anaphor are a synonymy pair in SynLex, and one feature consisting of the SynLex score for that word pair (if there is one). For example, the word *företag* ('business') has three synonyms in SynLex, with scores ranging from 3.2 to 4.0:

4.0 firma ('firm')

3.3 bolag ('corporation', 'company')

3.2 affärsverksamhet ('business (activity)')

and the 4.0 synonym *firma*('firm') is in turn listed with four synonyms:

4.4 rörelse ('enterprise')

4.0 företag ('business')

3.1 bolag ('corporation', 'company')

3.1 affärsverksamhet ('business (activity)')

Thus, the word pair *företag* and *bolag* would get a SynLex score of 3.3 in addition to a positive binary feature, whereas the word pair *företag* and *rörelse* would get a SynLex score of 0.0 and a negative binary feature.

3.2 Semantic relatedness in word-space models

Since the early 90's, a large body of research has developed which aims at capturing (lexical) semantic meaning through analyzing word co-occurrence and distribution (Grefenstette, 1994; Schütze, 1998). In analogy with the strongly

[2] Any influence on classification is likely to transfer to the complete coreference chains.

[3] URL: http://lexikon.nada.kth.se/synlex.html

related vector-space model, the representational models in these theories are commonly referred to as *word-space* models. Sahlgren (2006) argues that we can classify word-space models into two main groups: one which defines co-occurrence as two words occurring in the same document and one which defines it as two words occurring within a fixed-size sliding window. The first type is claimed to capture *syntagmatic* relations between words, the second type instead captures *paradigmatic* relations. Sahlgren (2006) gives credence to these claims through a series of experiments, but also shows that there is quite a bit of overlap between the two types. We investigate the effectiveness of these two types of models, separately as well as in conjunction, on the current task, using the standard cosine similarity measure.

Many researchers have experimented with applying *singular value decomposition* (SVD) (Golub and van Loan, 1996) to the matrices used by the word-space models to store the co-occurrence data. This process can be used for a dimensionality reduction for the similarity vectors. When the objects represented by the matrix are words and documents, this procedure is often called *latent semantic analysis* (LSA) and it is described in (Deerwester et al., 1990) and given a psychological motivation in (Landauer and Dumais, 1997). The advocates of LSA claim that it allows for capturing "latent" relations among words, that are not accessible through the raw co-occurrence data. In addition to the similarities calculated from the unprocessed matrices, we therefore also examine the effects of using singular value decomposition on the two types of word-spaces described above (again using the cosine similarity measure).

3.2.1 Term selection techniques

Another closely related approach to capturing similarities between words are so-called term selection or term weighting techniques. Just like the word-space models, their modeling capabilities are based on co-occurrence analysis. Where word-space models are based in geometry, term selection techniques are based in statistics or information theory. We use the mutual information (MI) measure (also referred to in (Manning and Schütze, 1999) as expected mutual information) on the two types of co-occurrence mentioned previously (within document or within a sliding window) and compare the results on the current task.

3.2.2 Building the word-space models

The corpus used for training the word-space models comes from the same newspaper and domain as the coreference annotated data (described in Sect. 4.1). It consists of about 1.5 million running words. When training the word-space model, we also include the coreference annotated data in the training data. However, this is not a case of "testing on the training data", since the annotations in the coreference data are not taken into consideration by the word-space model. The word-space model needs to see the words it is modeling as they occur in running text, and the more such examples provided, the better the model will function, typically. The coreference annotated data is just treated as another source for collecting co-occurrence data by the word-space model; the coreference data does not constitute a gold standard for this part of our system.

3.2.3 Word-space features

We thus have three models of similarity: using cosine or mutual information on vectors from the co-occurrence matrices (we merely apply a standard log-2 frequency damping) or using cosine on the dimensionality reduced vectors.

Table 1 gives an overview of all the word-space features, and the three models are represented by the three rows in the table. Each of these three models has two variants: the context window-based (column 'a') and the document-based (column 'b'). The score for the head words of each candidate anaphor-antecedent pair from each model is used as a feature, describing to what degree the two NPs are related within the respective models. We also extract a binary feature for each model, which is positive only for the highest-ranking coreference candidate for each NP within a document (columns 'c' and 'd'). Finally, we create sets consisting of the top 10 most similar coreference candidates for every definite description and proper noun within a document. This is done for each model and similarity measure, with one set containing context window-based (column 'e') and one set containing document-based relations (column 'f'; see also Fig. 1). At least when using the cosine measure on the non-reduced vectors,[4] we are hoping that these sets will help us distinguish between words that are syntagmatically

[4]We do not rule out the same effect for the MI measure or for the SVD-reduced matrix, but it has only been demonstrated for the non-reduced vectors and the cosine measure.

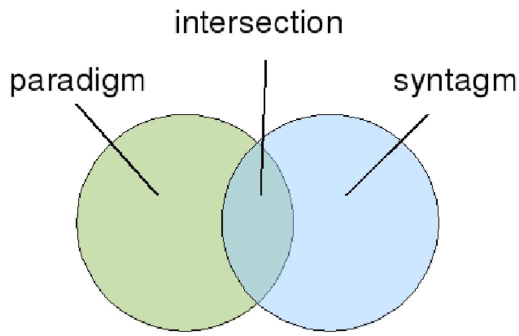

Figure 1: Forming three sets of words: paradigmatically related (window-based co-occurrence), syntagmatically related (document-based co-occurrence) and the intersection of these.

and paradigmatically related. We create a binary feature for each of the three sets formed this way (column 'g' represents the intersection of the previous two), hopefully indicating the type of relation (or lack thereof) in which a particular pair of words stand.

4 Classification of Pairs of Definite Descriptions

4.1 Data

The annotation of the data used in these experiments was done by one of the authors, based on the BREDT annotation guidelines for referential relations developed for Norwegian (Borthen, 2004) with minor modifications for the language (Swedish) and the domain (economic news text). The main goal of the annotation is to mark a select set of anaphoric and cataphoric relations. The most frequent, and thus the most important one, is coreference, which is defined as a relation of *identity of reference*. The annotated data we use here consists of 66 documents; there is a total of 6606 noun phrases of which 1887 (28%) are annotated as coreferent.

The preprocessing includes part-of-speech tagging and lemmatization with Granska (Carlberger and Kann, 1999), dependency parsing with MaltParser (Nivre et al., 2007), Named Entity tagging, and NP chunking. For NEs, basic semantic information is added by extending each occurrence of the NE type 'organization' with the synset for *företag, organisation* ('company', 'organization'), and the NE type 'person' with the synset for *människa, person* ('human being', 'person') from the online version of the Swedish WordNet[5] (Viberg et al., 2002).

Since we define coreference as a relation of identity of reference, each NP within a coreference chain is coreferent with all other NPs within that chain. Thus, in order to construct pairs of anaphors and candidate antecedents, each NP is combined with all other NPs within the document. As stated in Sect. 3, we are concerned with two sub-tasks in these experiments; for the first task there are 269 positive instances, and for the second 328. The data is partitioned so that the instances used in the two experiments are disjoint.

4.2 Features

Our baseline feature set is comprised of language- and domain-independent features used in high-performing coreference resolution systems such as (Soon et al., 2001) and (Strube et al., 2002), some domain-dependent features handling e.g., quoted speech, and some features based on corpus studies on definite descriptions by e.g., (Fraurud, 1992) and (Vieira and Poesio, 2000) describing e.g., NP complexity. This feature set includes 90 features; 58 of these features describe each NP in a candidate anaphor-antecedent pair (including gender, number and definiteness, as well as syntactic function and approximations to salience), and 32 features describe the candidate pair in terms of morphological similarity and syntactic parallelism, location (e.g., whether the two NPs are located within the same sentence, or in adjacent ones), and string similarity (e.g., complete and partial overlap, and the Levenshtein distance). Classification with this feature set is used as a baseline.

In addition to this standard feature set, semantic information is added via two SynLex features (described in section 3.1), and 21 word-space features (described in section 3.2). We group the 21 word-space features into six different configurations as such (please also refer to Table 1):

- **WS**: includes all 21 word-space features (WS stands for word-space)
- **WS cosine**: all features in row 1 in Table 1
- **WS MI**: all features in row 2
- **WS SVD**: all features in row 3 (we use a standard dimensionality of 200 in our experiments)

[5]URL: http://www.lingfil.uu.se/ling/swn.html; We do not at present have access to SWN in a machine readable format.

	window	document	window*	document*	paradigm*	syntagm*	intersection*
cosine	1a	1b	1c	1d	1e	1f	1g
MI	2a	2b	2c	2d	2e	2f	2g
SVD	3a	3b	3c	3d	3e	3f	3g

Table 1: Features from word-space models. The *-character indicates that features in the marked column are binary. 'MI' stands for mutual information and 'SVD' for singular value decomposition.

- **WS window**: all features in columns a, c, and e–g; aims to capture paradigmatic relations (we use a standard window size of 3 words to each side of the focus word in our experiments)
- **WS document**: all features in columns b and d–g; aims to capture syntagmatic relations

4.3 Classification

For classification of pairs of definite descriptions, we use 5-fold cross validation with the memory-based learner TiMBL (Daelemans and van den Bosch, 2005). We use the IB1 (k-nn) algorithm with k=5, the distance metric MVDM/overlap, and gain ratio feature weighting, and feature sets adapted for each task.

The classification is evaluated on instance level using the following measures: *precision*, *recall*, and *F-score*. Precision is defined as the number of correct coreference relations given by TiMBL divided by the total number of coreference relations given by the system. Recall is the number of correct coreference relations given by TiMBL divided by the total number of coreference relations in the data. F-score is the harmonic mean of precision and recall.

5 Results

The results in Tables 2 and 3, below, show a positive effect from the semantic features, though not in all configurations. The SynLex features do not provide any useful information to the system – their only effect is to lower the recall slightly. One might argue that the comparison between SynLex and our word-space models is unfair, as SynLex is a general resource whereas the word-space models are domain-specific. But this is in fact the point we wish to make: in order to handle coreference between noun phrases, we need domain-specific models of semantic relatedness. All but one configuration of word-space features produce higher precision than the baseline feature set, and the majority also give a simultaneous increase in recall.

	Precision	Recall	F-score
Baseline	28.3	22.1	24.8
SL and WS	–	–	–
SL	–	–	–
WS	30.7	18.6	23.1
WS cosine	27.7	19.0	22.5
WS MI	33.7	**23.7**	27.8
WS SVD	32.9	22.1	26.5
WS window	31.9	20.9	25.3
WS document	**34.7**	**23.7**	**28.2**

Table 2: Micro-averaged results: antecedent is an NE, anaphor is a common noun. SL stands for 'SynLex'. The feature sets are named and described in Sect. 4.2, above. SynLex does not contain names, therefore we cannot calculate results for settings involving this resource.

	Precision	Recall	F-score
Baseline	42.7	9.8	15.9
SL and WS	48.3	8.8	14.9
SL	42.1	9.8	15.8
WS	49.1	8.8	15.0
WS cosine	**52.9**	11.0	18.2
WS MI	50.7	10.7	17.6
WS SVD	51.2	**12.5**	**20.1**
WS window	43.3	8.8	14.7
WS document	48.6	10.4	17.1

Table 3: Micro-averaged results, both antecedent and anaphor are common nouns.

For the data set where both antecedent and anaphor are common nouns (set '2a' in Sect. 1), we see that the word-space model where we have applied SVD gives the best results, though the "raw" model actually gives higher precision (Table 3). This is not too surprising; given that the SVD is applied in order to uncover latent relations, we can expect a high recall – at the cost of a certain level of noise creeping in, resulting in a lower precision than for the "raw" model.

More surprising was to see that the models with co-occurrence being defined on a document level give better results on both tasks than the ones where it is based on the sliding context windows. We expected the latter to capture paradigmatic relations better than the former, but other factors, perhaps related to data sparseness, seem to influence the results contrary to our intuition. It can be argued, however, that the SVD can manage to capture paradigmatic information even when considering co-occurrence on a document level (features 3b and 3d – 3g in Table 1); that this in fact constitutes part of the "latency" in LSA. Further, in the task where the antecedent is an NE, it may well be that the relation between the two NPs is better thought of as syntagmatic than paradigmatic.

We also see that the 'WS MI' feature setting performs well on the task where the antecedent is an NE. It has been argued (Manning and Schütze, 1999) that the MI measure favors rare cases; something which applies to the NEs, and therefore could explain why this feature setting does well on this task.

The subtask where both antecedent and anaphor are common nouns can conceptually be split further into two cases. First, we have cases that can be resolved using information on lexical relations between the head nouns of the anaphor and the candidate antecedent; relations such as (near) synonymy, as in 'the business' - 'the company', or hypernymy, as in 'mediator' - 'the profession'. Second, we have cases that require additional information for resolution, e.g., common-sense reasoning or real-world knowledge as in 'the period April-June' - 'the second quarter', and/or keeping track of the current focus 'two metal workers' - 'the dismissed (employees)'. We expect the word-space approach to deal better with the former cases than the latter, but we cannot exclude that the latter, too, will display some degree of similarity in a word-space model.

We performed an experiment where we used the word-space features exclusively (no baseline features were used) for classifying the instances. This results in rather low figures in terms of precision and recall, but the successful cases may still give us an idea of the type of information we can hope to extract. E.g., the word-space models correctly predicted a coreference relation between *siffror* and *statistik* ('numbers' and 'statistics'), *anställda* and *personal* ('employees' and 'personel'), and *euroområdet* and *euroländerna* ('the Euro area' and 'the Euro countries'). These are all cases of near synonymy, and the results thus support our assumption that the word-space model will handle such cases better than cases where focus or reasoning play a part in the resolution.

We have performed these experiments on Swedish news text, but we have reasons to believe that the results are at least partly generalizable. First of all, the problem of having to resolve non-identical head definite descriptions exists and is relevant for other languages than Swedish, as we discussed in Sect. 1. Secondly, word-space models can be constructed for any language and domain where the tokenization of text into words is not a major issue. Finally, though they do not employ word-space features directly, Hendrickx et al. (2008b) and Ng (2007) show, for Dutch and English, that including semantics from statistically based corpus-methods has positive effects on the accuracy on their systems.

6 Conclusion

Coreference resolution of definite NPs is a complex problem, resulting in higher error rates compared to Named Entity coreference resolution, or pronoun resolution. One reason for this is the problem of acquiring various types of domain-specific lexico-semantic and common-sense knowledge needed for resolution. We present encouraging results from a study on using word-space similarity measures to approximate this knowledge in a system for resolution of definite descriptions.

References

Kaja Borthen. 2004. Annotation scheme for BREDT. Version 1.0. Technical report, University of Bergen.

Johan Carlberger and Viggo Kann. 1999. Implementing an efficient part-of-speech tagger. *Software Practice and Experience*, 29:815–832.

Walter Daelemans and Antal van den Bosch. 2005. *Memory-Based Language Processing*. Studies in Natural Language Processing. Cambridge University Press.

Scott Deerwester, Susan Dumais, Thomas Landauer, George Furnas, and Richard Harshman. 1990. Indexing by Latent Semantic Analysis. *Journal of the American Society of Information Science*, 41(6):391–407.

Christiane D. Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.

Kari Fraurud. 1992. *Processing Noun Phrases in Natural Language Discourse*. Ph.D. thesis, Stockholm University.

Daniela Goecke, Maik Stührenberg, and Tonio Wandmacher. 2007. Extraction and representation of semantic relations for resolving definite descriptions. In *OTT'06. Ontologies in Text Technology: Approaches to Extract Semantic Knowledge from Structured Information. Publications of the Institute of Cognitive Science (PICS) 1-2007*, Osnabrück, Germany.

Gene H. Golub and Charles F. van Loan. 1996. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, USA, 3 edition.

Gregory Grefenstette. 1994. *Explorations in Automatic Thesaurus Discovery*. Kluwer Academic Publishers, Boston, MA, USA.

Iris Hendrickx, Gosse Bouma, Frederik Coppens, Walter Daelemans, Veronique Hoste, Geert Kloosterman, Anne-Marie Mineur, Joeri Van Der Vloet, and Jean-Luc Verschelde. 2008a. A coreference corpus and resolution system for Dutch. In European Language Resources Association (ELRA), editor, *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, May.

Iris Hendrickx, Véronique Hoste, and Walter Daelemans. 2008b. Semantic and Syntactic Features for Dutch Anaphora Resolution. In *Proceedings of the 9th International Conference on Intelligent Text Processing and Computational Linguistics, CICLing 2008, Haifa, Israel, February 17-23, 2008*, Lecture Notes in Computer Science. Springer.

Viggo Kann and Magnus Rosell. 2006. Free construction of a Swedish dictionary of synonyms. In S. Werner, editor, *Proceedings of the 15th NODALIDA conference, Joensuu 2005*, Ling@JoY : University of Joensuu electronic publications in linguistics and language technology 1, Joensuu. SBN 952-458-771-8, ISSN 1796-1114.

Thomas Landauer and Susan Dumais. 1997. A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, 104(2):211–240.

Christopher Manning and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, MA.

Ruslan Mitkov. 2003. Anaphora Resolution. In Ruslan Mitkov, editor, *The Oxford Handbook of Computational Linguistics*, pages 266–283. Oxford University Press.

Thomas Morton. 2005. *Using Semantic Relations to Improve Information Retrieval*. Ph.D. thesis, University of Pennsylvania.

Vincent Ng. 2007. Semantic class induction and coreference resolution. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 536–543, Prague, Czech Republic, June. ACL.

Joakim Nivre, Johan Hall, Jens Nilsson, Atanas Chanev, Gülşen Eryiğit, Sandra Kübler, Svetoslav Marinov, and Erwin Marsi. 2007. Maltparser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, 13(2):95–135.

Massimo Poesio, Sabine Schulte im Walde, and Chris Brew. 1998. Lexical Clustering and Definite Description Interpretation. In *Proceedings of the AAAI Spring Symposium on Learning for Discourse*, Stanford, CA, March.

Magnus Sahlgren. 2006. *The Word-Space Model: Using Distributional Analysis to Represent Syntagmatic and Paradigmatic Relations between Words in High-Dimensional Vector Spaces*. Ph.D. thesis, Stockholm University, Stockholm, Sweden.

Hinrich Schütze. 1998. Automatic Word Sense Discrimination. *Computational Linguistics*, 24(1):97–123.

Wee Meng Soon, Hwee Tou Ng, and Daniel Chung Yong Lim. 2001. A Machine Learning Approach to Coreference Resolution of Noun Phrases. *Computational Linguistics*, 27(4):521–544.

Michael Strube, Stefan Rapp, and Christoph Müller. 2002. The influence of minimum edit distance on reference resolution. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 312–319, Philadelphia, PA, USA, July. ACL.

Åke Viberg, Karin Lindmark, Ann Lindvall, and Ingmarie Mellenius. 2002. The Swedish WordNet Project. In *Proceedings of Euralex 2002, Copenhagen University*, pages 407–412.

Renata Vieira and Massimo Poesio. 2000. An empirically based system for processing definite descriptions. *Computational Linguistics*, 26(4):539–593.

Piek Vossen. 1998. Introduction to EuroWordNet. *Computers and the Humanities*, 32(2–3), March.

ISSN 1736-6305 Vol. 4
http://hdl.handle.net/10062/9206