

# Analysis and Reference Resolution of Bridge Anaphora across Different Text Genres

Iris Hendrickx<sup>1</sup>, Orphee De Clercq<sup>2</sup> and Veronique Hoste<sup>2,3</sup>

<sup>1</sup> Centro de Linguística da Universidade de Lisboa, Lisboa, Portugal  
`iris@clul.ul.pt`

<sup>2</sup> LT3, School of Translation Studies, University College Ghent, Belgium  
`orphee.declercq@hogent.be`, `veronique.hoste@hogent.be`

<sup>3</sup> Faculty of Linguistics, Ghent University, Ghent, Belgium.

**Abstract.** We discuss bridge relations in Dutch between two textual referents across six different text genres. After briefly presenting the annotation guidelines and inter-annotation agreement results, we conduct an in-depth manual analysis of the different types of bridge relations found in our data sets. This analysis reveals that for all genres bridging references stand mostly in a class relationship, which is exactly the kind of information represented in a WordNet hierarchy. This inspired us to investigate to what extent a standard coreference resolution system for Dutch is capable of resolving bridge relations across different text genres and study the effect of adding semantic features encoding WordNet information. Our results reveal modest improvements when using Dutch WordNet LCS information for all but one genre.

## 1 Introduction

Automatic coreference resolution systems are defined as systems that automatically detect references to discourse entities in a text. Often the same entity is mentioned multiple times in a text and it is the task of the resolution system to determine which mentions refer to the same entity (we will denote these as “identity” relations) and which ones introduce a new entity in the discourse. Besides these clear cases of direct references to previously mentioned entities, texts can also contain indirect, or more vague relations between two mentions. One of the earlier studies to introduce the term *bridging* for these instances was conducted by Clark [2]. He documents several scenarios in which an inference step is needed to grasp the meaning intended by the speaker. An important conclusion Clark draws about bridge relations is that the language or text itself does not offer the solution to resolve this inference relation, the listener or reader needs to use its world knowledge about the anaphor and antecedent to make the correct interpretation. Moreover, Clark also lists several types of bridging such as set-membership, part-whole relations, roles, reasons and consequences.

In this paper we discuss the annotation and resolution of bridge relations in Dutch across six different text genres: administrative texts (ADM) such as notes from political meetings or official speeches, texts

used for external communication (EXT) such as commercial brochures, instructive texts (INST) for example instruction manuals of domestic products, newspaper text (JOUR), medical encyclopedic texts (MED) and wikipedia text (WIKI). The broad definition for annotating bridge relations in the guidelines followed [1], and low inter-annotator agreement results necessitated an in-depth analysis of what had exactly been annotated as *bridge* in our data. In order to do so we decided to manually analyze 50 examples of each genre following a recently developed typology suited for bridging relation annotation [21]. Based on this analysis and previous studies on resolving bridge relations (e.g. [16, 15, 10]), we also present a first version of an automatic resolution system for Dutch bridge relations. We wish to investigate to what extent a standard mention-pair coreference resolver is applicable to the task, how it behaves across different genres, and study the usefulness of WordNet (WN) for resolving bridge relations in our data sets since computing semantic distance based on WN has proven a useful feature in previous work [15].

## 2 Related Work

Bridging anaphora have been extensively investigated by Poesio et al. [17, 16, 25, 15, 10]. Both the typology of bridging references and different information sources to improve automatic resolution were studied. Since bridging reference resolution often requires some form of world knowledge – or at least information that is not explicitly represented in the textual context of the reference – several investigations were conducted to measure the usefulness of WordNet as an external resource to resolve bridge relations. It was shown, however, that information extracted from WordNet does not necessarily help resolving bridge relations [25]. Another method is to use syntactic patterns that are able to express part-whole relations such as “the x of the y”. If such patterns of a potential antecedent and anaphor are found with a high frequency in a corpus [14] or on the web [15, 10], this provides a strong clue that they could stand in a bridge relation. Besides the above-mentioned features, in [15] features expressing salience, such as utterance distance and first-mention were also studied. Here, WordNet and Google were employed to compute lexical distances between referents and the study revealed that using the hypernym relations in WordNet gives comparable performance to using Google. [20] and [19] study bridge relations from a more theoretical viewpoint and focus on how to distinguish bridge relations from identity relations in the annotation task. This will be discussed in closer detail in Section 3.1.

## 3 Annotation of Bridge relations

For this study we use bridge relations that have been annotated during two Dutch corpus projects: SoNaR [23]<sup>4</sup> and COREA [4]<sup>5</sup>. In both

---

<sup>4</sup> SoNaR is currently still under development.

<sup>5</sup> These data sets are available at: <http://www.inl.nl/tst-centrale>

**Table 1.** Statistics on the 6 different data sets and the type of annotated referential relations.

genre	#texts	#tokens	IDENT	BRIDGE	PRED	BOUND
ADM	41	49,335	3,547	80	242	5
EXT	129	149,501	11,066	383	519	31
INST	26	38,001	3,869	92	1,566	2
JOUR	172	127,040	10,463	318	495	67
MED	499	74,445	4,924	1,779	289	19
WIKI	66	177,766	16,533	504	1,007	6

projects different text types were annotated with coreferential relations. In the present study we decided not to use all available annotated data, more informal text genres such as blogs or spoken texts were excluded since these are difficult to process automatically. Six different text genres were selected, more information and some data statistics are presented in Table 3. Besides bridge relations, who are at the subject of this study, the coreference annotation includes identity, predicative and bound (e.g. ‘every man has his problems’) relations. The data sets differ widely in size, document length and annotated relations: INST is the smallest genre, both in the number of documents and tokens. In comparison to the other genres, the MED data set contains a remarkably high amount of annotated bridge relations, which can be explained by this genre’s explanatory nature. The data consists of short encyclopedic descriptions; a human body part or a disease, for example, is first described in general terms followed by more specific features

In both projects coreference was annotated according to the same COREA guidelines [1]. In these guidelines a bridge relation is vaguely defined as ‘a partial coreference relation exists when reference is made to a subpart of an object that has already been mentioned in the discourse’, after which some Dutch examples are presented that demonstrate two types of bridge relations: part-whole and subset/superset.

Here are some examples of how bridge relations have been annotated. Example 1 is taken from the JOUR data set and illustrates a superset-subset relation in which the antecedent expresses a rather general class whereas the anaphoric element is more specific. Example 2, also taken from a newspaper article, represents a part-whole relation and Example 3 (WIKI) shows two entities almost referring to the same thing. The first referent focuses on a published series whereas the second one refers to the physical albums.

*Example 1.* **Natuurlijke therapieën** zoals **voedingstherapie**, toepassing van warmte, lucht, licht en water. (Eng: **Natural therapies** such as **nutritional therapy**, application of heat, air, light and water.)

*Example 2.* Voor **de website** hebben we de laatste tijd veel inspanningen gedaan . Van op **de Engelstalige homepage** kan je nu bijvoorbeeld met één klik naar een volledig aanbod van de Engelstalige masterprogramma’s. (Eng: A lot of effort has been put into **the website** lately.)

**Table 2.** Inter-annotator agreement scores for bridge relations using the COREA-guidelines

genre	#tokens	#bridges	$F_{\beta=1}$	
			total match	
ADM	1548	17	3	31.25
EXT	1134	24	13	71.57
INST	3631	40	5	24.24
JOUR	2796	46	10	39.28
MED	2260	112	44	60.08
WIKI	2444	79	22	51.42
Total	13813	318	97	46.30

*With only one click you can, for example, go from **the English homepage** to the full range of English Master’s programs.)*

*Example 3.* De eerste verhalen werden gepubliceerd in wat **de Rode reeks** wordt genoemd ( zo genoemd omdat **de albums** een helrode omslag hadden, die reeks heel herkenbaar maakte). (*Eng: The first stories were published in what is called **the Red series** (given this name because **the albums** had a bright red color which made the series easily recognizable).*)

Other projects involving bridge annotation usually use a more fine-grained typology to label different types of bridge relations. In the GNOME corpus [13] bridge relations like set membership, subset, and generalized possession (i.e. part-whole) are annotated. The Prague Treebank [11] includes extended coreference annotation of bridge relations: part-whole, subset, functional, but two additional groups (“contrast” and “other” ) are also annotated in order to capture all other types of bridge relations, such as people-location or event-argument relations. The Copenhagen Dependency Treebank [8] has a very detailed annotation scheme for associative anaphora based on Pustejovsky’s qualias and general semantic roles. It includes, for example, annotation of agent relations between two phrases like “the operation” and “the surgeon”.

Since the COREA guidelines are not very specific nor detailed in their description of bridge relations and marking bridging references is a difficult task (cf. [18]), we can expect low inter-annotator agreement (IAA). In order to test this, two linguists annotated a set of ten randomly selected texts, containing 13,813 tokens in total. Since the COREA and SoNaR corpora sample comprise texts from six different domains at least one text of each genre was annotated. As evaluation metric we computed F-scores ( $\beta = 1$ ) by each time taking one linguist as gold standard and scoring the annotations of the other for precision and recall. The results per genre can be found in Table 2, in which token counts, the total number of bridges annotated by either one of the annotators and the number of bridges on which both annotators agree are also included.

As expected, the results reveal low agreement scores for all genres, on average 46%, and most notably instructive texts with an IAA of 24% F-score. The level of complexity of the INST texts, containing many technical terms, is one of the main causes for this low agreement. When we compare this with other IAA scores on this type of task, for example [16] and [18], we find similar low agreement scores. Having a closer look at those bridging references on which both annotators do agree we note that these stand mostly in a subset/superset relation and disagreements are mostly due to interpretation differences and inconsistent annotation. Example 4 (from the JOUR data set) aptly illustrates how difficult the annotation task can be. Determining the meaning of this sentence requires very detailed and specific world knowledge. First of all, you need to know that a “V-twin” is a specific type of engine. Making the correct interpretation requires information about whether all Harley Davidson motors are equipped with this type of engine or not. Two possible interpretations exist here: if each Harley has this engine, the example expresses a metaphoric relation in which “V-twin” refers to all Harleys in general. However, it could also be the case that the writer merely sketches an image by referring to a specific type of Harley, a heavy V-twin, in that case there would be a bridge relation between Harley and V-twin<sup>6</sup>.

*Example 4.* Maar het beeld van **de Harley** beklijft wel: onderuitgezakt zitten op **een zware “V-twin”**, met scènes uit de film “Easy Rider” in je hoofd. *Eng: But the image of the Harley is enthralling: slouched on a heavy “V-twin”, with scenes from the movie “Easy Rider” going through your head.*

### 3.1 Analysis

Since the guidelines followed do not accurately define bridge relations and the IAA results are rather low, we decided to analyse the annotations in detail. With this analysis we aimed to gain better insights into our data, i.e. to know what was exactly denoted as a bridge relation and to better understand what type of information is needed to resolve these automatically. To this purpose 50 Dutch examples were randomly selected from each of the different domains after which we tried to categorize these into the more fine-grained typology proposed by Recasens et al. [20]. They discuss a typology of *near-identity* relations closely resembling bridge relations and divide this type of relation into four broad classes: metonymy, meronymy, class type and spatio-temporal functions, each containing several subtypes. In [19] an even broader definition of bridge relation is presented, here referential relations are perceived as a continuum between identity and non-identity relations. Two different types of near-identity operations are identified: *neutralization* captures cases in which referents are considered to be equal and where distinctive attributes are ignored as one concentrates on what is in common. The other operation is *refocusing* which applies to referents that have a more specific or different attribute in focus than its antecedent. As

---

<sup>6</sup> In fact, all Harley-Davidson motors have a V-twin engine.

**Table 3.** Manual analysis according to the typology for Near Identity.

Type	subtype	MED	WIKI	JOUR	ADM	EXT	INST
Meronymy	Part-Whole	6	9	13	1	3	2
Meronymy	Stuff-Object	0	0	0	0	2	0
Meronymy	Set-Set	5	11	11	7	12	2
Class	More specific	32	24	18	20	32	40
Class	More general	2	2	5	14	2	3
Other		4	4	3	4	0	3

the division in [20] offers a systematic annotation scheme, we decided to adopt this typology for our analysis. As far as the near-identity typology is concerned, we are mostly interested in the meronymy and class types since these match our definition of bridge relations in the COREA guidelines. For example in 1 we see a typical class type bridge relation in which the anaphor *voedingstherapie* is clearly a type of *Natuurlijke therapieën*. We were unable to completely follow the proposed typology completely because it also includes some additional classes that were not annotated as bridge-identity in our data; we chose to annotate metonymy as identity relations and we have special separate markers to denote the spatio-temporal different references and modality and negation.

The left side of table 3 lists the types and subtypes of the typology that we did use. The broad class meronymy is divided into part-whole relations where one entity is a functional part of the whole entity, stuff-object relations that cover the “is made of” relations and set-set relations that denote two (vaguely defined) overlapping sets. An example of a set-set relation can be found in Ex. 3.

We would like to stress that this type of annotation is highly subjective because language usage can lead to creative entity descriptions that are difficult to classify into nicely organized, predefined separate labels. This is also in line with the findings presented in [18] where 6 annotators were asked to classify difficult cases into the typology categories: for many sentences the annotators made different choices. For this reason, we added one additional type to the typology: all cases that do not fit the typology, that are erroneously labeled or that are highly ambiguous were labeled as “Other”.

The results of our manual analysis of 50 annotated bridge relations for each of the 6 different text genres can be found in Table 3. The most frequent type of bridge relation found for all genres is a class relationship in which the anaphor provides a more specific description than the antecedent (such as Example 1). These more-spec class relations make up 50-70 % of the cases. This implies that people tend to first introduce a general class, and then introduce a more specific entity referring back to this group. This tendency is particularly strong in the medical data, which can be explained by the explanatory content of these texts. For the ADM data set, however, we observe a remarkably high number of

relations going in the opposite direction, i.e. a specific entity is introduced first, and later a reference to a larger class. This data set contains transcribed speeches and in Example 5, taken from a transcription of a discussion in the Belgian Senate, this relation is illustrated. We suspect that this might be a rhetorical method to draw the audience’s attention.

*Example 5.* In het artikel 556 van hetzelfde wetboek worden zij ” die (... ) gestraft met **een geldboete van 10 tot 20 frank.** (...) **De straffen** waarin voorzien wordt lijken bovendien uiterst licht.

(*Eng: In article 446 from the same Code they who (...) will be punished with a fine of 10 to 20 francs. (...) The punishments provided seem extraordinarily light.*)

The second most common relation are overlapping sets (cf. Example 3), ranging from 10 to 25% for the different data sets. Part-whole relations seem rather scarce as they only occur in 1-10 % of the cases. This analysis thus reveals that although the annotated bridge relations do cover a quite diverse group of relations, the majority of the bridge relations concerns class type relations. Since this is exactly the type of information that can be extracted from the WordNet hierarchy, we decided to add additional WordNet features to our automatic coreference resolution system as will be explained in the next section.

## 4 Experiments

For our experiments we adopt a standard pairwise machine learning approach of Soon et al [24]. We view bridge relations, however, as a different kind of relations than the typical ‘identity’ relations. A common approach to automatically resolve coreferential identity relations is to first detect all referents in the text and make pairwise decisions between a potential anaphor and antecedent. As noted in Ng’s survey on coreference resolution [12] this approach has the limitation that classifications are made independent from each other and that an additional clustering step is always needed to resolve overlapping and contradicting individual pair-predictions. In a bridge relation this is not the case: elements are not equal and do not all refer to the same entity, moreover, each element has a clear-cut role in the relation. One referent, for instance, expresses a more generic concept than the other, or one referent is a functional part of the other referent. This implies that the pairwise approach is well suited to accommodate this type of relation.

Since bridge relations occur less frequently than identity relations, making negative instances for all other NP pairs leads to a very sparse data set even when our search space is restricted to the 20 previous sentences. Consequently, we decided to use a sampling approach in order to reduce the number of negative instances to a ratio of 1: 3 positive/negative instances which was also done by Poesio et al. in their study of bridge relations [15]<sup>7</sup>. Because of this sampling approach the size of the different data sets ranges from only 68 positive instances for the ADM data

---

<sup>7</sup> We would like to stress that this is a preliminary experimental set-up and are aware of the methodological issues that arise when both train and test data are sampled.

set to 1698 for the MED data (this is represented in the second column of Table 4). When we compare this with the actual amount of bridge relations available in our annotated data (see Table 3), we see that the counts are slightly lower which can be explained by the limited search space. For some bridge relations the anaphor and antecedent are more than 20 sentences apart. We did manage, however, to retrieve 78% of all bridges annotated in WIKI up to 95% in the MED data set.

We run two experiments: one with a standard feature set largely following the feature set of Soon et al, and one in which 4 additional features representing WN information were added. The coreference system and the features are described in more detail in [6, 5]. Since our manual analysis revealed that most bridge relations express some kind of general/specific relation between both referents which is well expressed in the WordNet hierarchy, we chose to focus on *least common subsumer* (LCS) information of WN hyperonym relations. To determine the LCS of two words, you follow their edges upwards with the shortest possible path in the WordNet hierarchy until you reach a concept (a common subsumer) that both words share. The LCS is represented as the number of edges (distance) between the two lemmas of the head words of the two candidate referents. We also used LCS to compute three different similarity metrics: Resnik's similarity measure [22], Jiang & Conrath's similarity measure [7] and Lin's similarity measure [9]. As wordnet we use the Cornetto database<sup>8</sup>, a lexical-semantic database for Dutch which combines EuroWordNet for Dutch and Referentiebestand Nederlands (RBN).

As our ML classifier we use Timbl [3], a k-nearest neighbor algorithm with default parameter settings. We each time perform 10-fold cross validation experiments and compute overall accuracy, and precision, recall and F-score for the bridge relations.

## 5 Results

Results are presented in Table 4, on the left side the results from the standard coreference resolution system are shown and on the right side the version that uses four additional WN features. In general the referential resolution system is able to correctly predict bridge relations with an F-score between 42.18% for the ADM data (without WN) and 65.69% for the JOUR data set (with WN). In general we observe that recall scores are clearly higher than precision and notable differences in performance are observed between the data sets individually - JOUR, WIKI and MED score better than ADM, EXT and INST. This can be partly explained by the higher number of annotated unambiguous bridge relations in the training data. When we compare the performance of the system with and without the WN features, we see that these features do increase performance, most noticeably for the ADM data set (+ 6.58%). We also observe that in general the recall scores drops and precision goes up when WN features are added. Only for the INST data set no improvement can be perceived at all (- 0.97%). We inspected the errors

---

<sup>8</sup> Available at <http://www.inl.nl/nl/lexica/cornetto>

**Table 4.** Results on bridge relation resolution

genre	# bridges	without WN				with WN			
		acc	recall	prec	$F_{\beta=1}$	acc	recall	prec	$F_{\beta=1}$
ADM	68	41.54	85.29	34.79	42.18	54.41	86.76	33.91	48.76
EXT	347	58.43	75.79	34.79	47.69	60.73	74.35	36.13	48.63
INST	86	62.21	76.74	37.50	50.38	62.50	73.26	37.28	49.41
JOUR	290	75.69	87.93	50.80	64.39	77.84	84.83	53.59	65.69
MED	1698	71.52	76.27	45.86	57.28	72.67	74.79	47.11	57.81
WIKI	398	79.75	71.90	58.98	64.81	80.12	71.32	60.00	65.17

made by the system for these two outliers genres. For the ADM data set, containing the lowest number of bridge relations in training, we see that especially the recognition of true negatives increases because no similarity measures are found between two NPs which makes these features decisive. Looking at the INST data we notice that especially the type of bridge relations are far more complicated and therefore performance does not rise when adding WordNet features. A more fine-grained annotation is necessary to better understand these differences and to allow for a more in-depth error analysis.

## 6 Conclusion

We reported on the annotation and resolution of bridge relations in Dutch across six different text genres. In accordance with earlier research on bridging for other languages, we show low inter-annotator agreement scores, but also a great variability in the scores for the different genres, ranging from 24.2 % (INST) to 71.6% (EXT). A shallow analysis of these scores suggests that the complexity of the INST texts is the cause for the low agreement scores. Motivated by a more fine-grained annotation analysis which revealed that class relationships represent up to 70% of the annotated relations, we investigated whether adding WordNet LCS information improved the performance of an existing mention-pair model in revolving bridge relations. We showed modest performance improvements for five out of six text types when adding semantic information. This first attempt to resolve bridging relations in Dutch texts revealed some shortcomings with respect to the experimental setup and the vagueness of the existing annotation guidelines. Future work will include a more in-depth annotation of the data per bridge type, in order to determine which features aid the resolution of bridge relations and to allow for a fine-grained error analysis.

## 7 Acknowledgments

This work is funded by the STEVIN programme of the Dutch Language-Union within the framework of the SoNaR project under grant number

STE 07014 and the National Foundation for Science and Technology (FCT) under the project “PEst-OE/LIN/UI0214/2011.” Researcher under the Scientific Program *Ciência 2008*.

## References

1. G. Bouma, W. Daelemans, I. Hendrickx, V. Hoste, and A.-M. Mineur. The COREA-project, manual for the annotation of coreference in Dutch texts. Technical report, University Groningen, 2007.
2. H. Clark. Bridging. In *Proceedings of the Conference on Theoretical Issues in NLP*, pages 169–174, 1975.
3. W. Daelemans, J. Zavrel, K. Van der Sloot, and A. van den Bosch. TiMBL: Tilburg Memory Based Learner, version 6.3, Reference Guide. Technical Report ILK Research Group Technical Report Series no. 10-01, Tilburg University, 2010.
4. I. Hendrickx, G. Bouma, F. Coppens, W. Daelemans, V. Hoste, G. Kloosterman, A.-M. Mineur, J. V. D. Vloet, and J.-L. Verschelde. A coreference corpus and resolution system for Dutch. In *Proceedings of LREC 2008*, pages 144–149. Marrakech, Morocco, 2008.
5. I. Hendrickx, V. Hoste, and W. Daelemans. Evaluating hybrid versus data-driven coreference resolution. *Lecture Notes in Artificial Intelligence. Anaphora: Analysis, Algorithms and Application, Proceedings of DAARC 2007*, 4410:137–150, 2007.
6. V. Hoste. *Optimization Issues in Machine Learning of Coreference Resolution*. PhD thesis, Antwerp University, 2005.
7. J. Jiang and D. Conrath. Semantic similarity based on corpus statistics and lexical taxonomy. In *Proceedings of International Conference on Research in Computational Linguistics*, volume 33. Taiwan, 1997.
8. I. Korzen and M. Buch-Kromann. Anaphoric relations in the copenhagen dependency treebanks. In *Beyond Semantics: Corpus-based Investigations of Pragmatic and Discourse Phenomena Proceedings of the DGfS Workshop*, pages 83–98, 2011.
9. D. Lin. Automatic retrieval and clustering of similar words. In *Proceedings of the 17th international conference on Computational linguistics*, pages 768 – 774. Montreal, Quebec, Canada, 1998.
10. K. Markert, M. Nissim, and N. N. Modjeska. Using the web for nominal anaphora resolution. In *In EAACL Workshop on the Computational Treatment of Anaphora*, pages 39–46, 2003.
11. A. Nedoluzhko, J. Mírovský, and P. Pajas. The coding scheme for annotating extended nominal coreference and bridging anaphora in the prague dependency treebank. In *Proceedings of LAW III*, pages 108–111. ACL, Suntec, Singapore, August 2009.
12. V. Ng. Supervised Noun Phrase Coreference Research: The First Fifteen Years. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1396–1411, 2010.
13. M. Poesio. Discourse annotation and semantic annotation in the gnome corpus. In *Proc. of the ACL Workshop on Discourse Annotation*, 2004.

14. M. Poesio, T. Ishikawa, S. S. im Walde, and R. Vieira. Acquiring lexical knowledge for anaphora resolution. In *In Proceedings of the 3rd Conference on Language Resources and Evaluation (LREC)*, pages 1220–1224, 2002.
15. M. Poesio, R. Mehta, A. Maroudas, and J. Hitzeman. Learning to resolve bridging references. In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04)*, pages 143–150, 2004.
16. M. Poesio and R. Vieira. A corpus-based investigation of definite description use. *Computational Linguistics*, 24(2):183–216, 1998.
17. M. Poesio, R. Vieira, and S. Teufel. Resolving bridging references in unrestricted text. In *Proceedings of the ACL workshop on Robust Anaphora Resolution*, pages 1–6, 1997.
18. M. Recasens. *Coreference: Theory, Annotation, Resolution and Evaluation*. PhD thesis, Department of Linguistics, University of Barcelona, Barcelona, Spain, September 2010.
19. M. Recasens, E. Hovy, and M. Martí. Identity, non-identity, and near-identity: Addressing the complexity of coreference. *Lingua*, 121(6):1138 – 1152, 2011.
20. M. Recasens, E. Hovy, and M. A. Martí. A typology of near-identity relations for coreference (NIDENT). In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*. ELRA, Valletta, Malta, 2010.
21. M. Recasens, L. Márquez, E. Sapena, M. A. Martí, M. Tauleé, V. Hoste, M. Poesio, and Y. Versley. SemEval-2010 Task 1: Coreference resolution in multiple languages. In *Proceedings of the 5th International Workshop on Semantic Evaluations (SemEval-2010)*, pages 1–8. Uppsala, Sweden, 2010.
22. P. Resnik. Using Information Content to Evaluate Semantic Similarity in a Taxonomy. In *International Joint Conference on Artificial Intelligence*, volume 14, pages 448–453. Montreal, Canada, 1995. URL <http://www.cs.umbc.edu/~finin/pub/resnik.ps>.
23. I. Schuurman, V. Hoste, and P. Monachesi. Interacting Semantic Layers of Annotation in SoNaR, a Reference Corpus of Contemporary Written Dutch. In *Proceedings of LREC 2010*, pages 2471–2477. Valletta, Malta, 2010.
24. W. M. Soon, H. T. Ng, and D. C. Y. Lim. A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics*, 27(4):521–544, 2001.
25. R. Vieira and M. Poesio. An empirically based system for processing definite descriptions. *Computational Linguistics*, 26:539–593, 2000.