Are Morpho-Syntactic Features More Predictive for the Resolution of Noun Phrase Coordination Ambiguity than Lexico-Semantic Similarity Scores?

Ekaterina Buyko and **Udo Hahn**

Jena University
Language & Information Engineering (JULIE) Lab
Fürstengraben 30, 07743 Jena, Germany

`ekaterina.buyko|udo.hahn@uni-jena.de`

Abstract

Coordinations in noun phrases often pose the problem that elliptified parts have to be reconstructed for proper semantic interpretation. Unfortunately, the detection of coordinated heads and identification of elliptified elements notoriously lead to ambiguous reconstruction alternatives. While linguistic intuition suggests that semantic criteria might play an important, if not superior, role in disambiguating resolution alternatives, our experiments on the reannotated WSJ part of the Penn Treebank indicate that solely morpho-syntactic criteria are more predictive than solely lexico-semantic ones. We also found that the combination of both criteria does not yield any substantial improvement.

1 Introduction

Looking at noun phrases such as

'cat and dog owner'
'novels and travel books'

their proper coordination reading (and asymmetric distribution of coordinated heads) as

'cat owner' AND *'dog owner'*
'novels' AND *'travel books'*

seems to be licensed by the striking semantic similarity between *'cat'* and *'dog'*, and *'novels'* and *'books'*, respectively. If this were a general rule, then automatic procedures for the resolution of coordination ambiguities had to rely on the *a priori* provision of potentially large amounts of semantic

background knowledge to make this similarity explicit. Furthermore, any changes in languages or domains where such resources were missing (or, were incomplete) would severely hamper coordination analysis.

Indeed, previous research has gathered lot of evidence that conjoined elements tend to be semantically similar. The important role of semantic similarity criteria for properly sorting out conjuncts was first tested by Resnik (1999). He introduced an information-content-based similarity measure that uses WORDNET (Fellbaum, 1998) as a lexico-semantic resource and came up with the claim that semantic similarity is helpful to achieve higher coverage in coordination resolution for coordinated noun phrases of the form '*noun1* and *noun2 noun3*' than similarity measures based on morphological information only.

In a similar vein, Hogan (2007b) inspected WORDNET similarity and relatedness measures and investigated their role in conjunct identification. Her data reveals that several measures of semantic word similarity can indeed detect conjunct similarity. For the majority of these similarity measures, the differences between the mean similarity of coordinated elements and noncoordinated ones were statistically significant. However, it also became evident that these were only slight differences, and not all coordinated heads were semantically related as evidenced, e.g., by *'work'/'harmony'* in *'hard work and harmony'*. The significance tests did also not reveal particularly useful measures for conjunct identification.

Rus et al. (2002) in an earlier study presented an alternative heuristics-based approach to conjunct identification for coordinations of the form '*noun1* and *noun2 noun3*'. They exploit, e.g., look-ups in WORDNET for a compound noun as a con-

cept, and for the sibling relation between nouns in the coordination and report bracketing precision of 87.4% on 525 candidate coordinations. Although the authors demonstrated that WORDNET was really helpful in coordination resolution, the evaluation was only conducted on compound nouns extracted from WORDNET's noun hierachy and, furthermore, the senses of nouns were manually tagged in advance for the experiments.

Despite this preference for semantic criteria, one might still raise the question how far non-semantic criteria might guide the resolution of noun phrase coordination ambiguities, e.g., by means of the distribution of resolution alternatives in a large corpus or plain lexical or morpho-syntactic criteria. This idea has already been explored before by various researchers from different methodological angles including distribution-based statistical approaches (e.g., Chantree et al. (2005), Nakov and Hearst (2005)), similarity-based approaches incorporating orthographical, morpho-syntactic, and syntactic similarity criteria (e.g., Agarwal and Boggess (1992), Okumura and Muraki (1994)), as well as a combination of distribution information and syntactic criteria (Hogan, 2007a).

Statistical approaches enumerate all candidate conjuncts and calculate the respective likelihood according to a distribution estimated on a corpus. For the coordination *'movie and television industry'* the distributional similarity of *'movie'* and *'industry'* and the collocation frequencies of the pairs [*'movie'* - *'industry'*] and [*'television'* - *'industry'*] would be compared against each other. However, for such an approach only an F-measure under 50% was reported (Chantree et al., 2005). Unsupervised Web-distribution-based algorithms (Nakov and Hearst, 2005) achieved 80% on the disambiguation of coordinations of the fixed form *'noun1 and noun2 noun3'*. Hogan (2007a) presented a method for the disambiguation of noun phrase coordination by modelling two sources of information, *viz.* distribution-based similarity between conjuncts and the dependency between conjunct heads. This method was incorporated in Bikel's parsing model (Bikel, 2004) and achieved an increase in NP coordination dependency F-score from 69.9% to 73.8%.

Similarity-based approaches consider those elements of a coordination as conjuncts which are most 'similar' under syntactic, morphological, or even semantic aspects. Agarwal and Boggess

(1992) include in their NP coordination analysis syntactic and some semantic information about candidate conjuncts and achieve an accuracy boost up to 82%. Okumura and Muraki (1994) estimate the similarity of candidate conjuncts by means of a similarity function which incorporates syntactic, orthographical, and semantic information about the conjuncts. The model provides about 75% accuracy.

The resolution of coordination ambiguity can also be tried at parsing time. Charniak and Johnson (2005), e.g., supply a discriminative reranker that uses e.g., features to capture syntactic parallelism across conjuncts. The reranker achieves an F-score of 91%.

Recently, discriminative learning-based approaches were proposed, which exploit only lexical, morpho-syntactic features and the symmetry of conjuncts. Shimbo and Hara (2007) incorporate morpho-syntactic and symmetry features in a discriminative learning model and end up with 57% F-measure on the GENIA corpus (Ohta et al., 2002). Buyko et al. (2007) employ Conditional Random Fields (Lafferty et al., 2001) and successfully tested this technique in the biomedical domain for the identification and resolution of elliptified conjuncts. They evaluate on the GENIA corpus and report an F-score of 93% for the reconstruction of the elliptical conjuncts employing lexical and morpho-syntactic criteria only. At least two questions remain — whether the latter approach can achieve similar results in the newswire language domain (and is thus portable), and whether the incorporation of additional semantic criteria in this approach might boost the resolution rate, or not (and is thus possibly more parsimonious). The latter question is the main problem we deal with in this paper.

2 Data Sets for the Experiments

2.1 Coordination Annotation in the PENN TREEBANK

For our experiments, we used the WSJ part of the PENN TREEBANK (Marcus et al., 1993). Some researchers (e.g., Hogan (2007a)) had recently found several inconsistencies in its annotation of the bracketing of coordinations in NPs. These bugs were shown to pose problems for training and testing of coordination resolution and parsing tools. Fortunately, a re-annotated version has been provided by Vadas and Curran (2007), with a focus

on the internal structure of NPs. They added additional bracketing annotation for each noun phrase in the WSJ section of the PENN TREEBANK assuming a right-bracketing structure in NPs. In addition, they introduced tags, e.g., 'NML' for explicitly marking any left-branching constituents as in

(NP (NML (JJ *industrial*) (CC *and*) (NN *food*)) (NNS *goods*))

where '*industrial*' and '*food*' are conjuncts. In the example

(NP (DT *some*) (NN *food*) (CC *and*) (NN *household*) (NNS *goods*))

the structure of the noun phrase is already correct and should not be annotated further, since '*household goods*' is already right-most and is coordinated with '*food*'. Still, in the original PENN TREEBANK annotation, we find annotations of noun phrases such as

(NP (NN *royalty*) (CC *and*) (NP (NN *rock*) (NNS *stars*)))

that remain unchanged after the re-annotation process.

2.2 Coordination Corpus

We, first, extracted a set of 3,333 non-nested NP coordinations involving noun compounds and one conjunction, with a maximal number of nine nouns (no prepositional phrases were considered). We focused on two patterns in the re-annotated WSJ portion:

(1) Noun phrases containing at least two nouns and a conjunction as sister nodes as in

(NP (NML (NN *movie*) (CC *and*) (NN *book*)) (NNS *pirates*))

or in

(NP (DT *some*) (NN *food*) (CC *and*) (NN *household*) (NNS *goods*))

(2) Noun phrases containing at least two noun phrases and a conjunction as sister nodes (as they remained unchanged from the original PENN TREEBANK version). Thereby, the second noun phrase contains at least two nouns as sister nodes as in

(NP (NP (NNP *France*)) (CC *and*) (NP (NNP *Hong*) (NNP *Kong*)))

We removed from this original set NPs which could not be reduced to the following pattern:[1]

---

[1] These are typically coordinations of the form '$(W)N_1$ and N_2', e.g., '*government sources and lobbyists*', where W is a sequence of i tokens ($i \geq 0$). 646 coordinations of this type occurred in the WSJ portion of the PTB.

$(W) N_1$ and $(W) N_2 N_3$,

where (W) is a sequence of i tokens with $i \geq 0$ as in '*street lamps*$_{N_1}$ *and ficus*$_{N_2}$ *trees*$_{N_3}$'.

The remaining major data set (A) then contained 2,687 NP coordinations. A second data set (B) was formed, which is a proper subset of A and contained only those coordination structures that match the following pattern:

$(X) N_1$ and $(W) N_2 N_3$,

where (X) is defined as a sequence of i tokens ($i \geq 0$) with all part-of-speech (POS) tags except nouns and (W) defined as above; e.g., '*a happy cat*$_{N_1}$ *and dog*$_{N_2}$ *owner*$_{N_3}$'. Test set B contains, in our opinion, a selection of less 'hard' coordinations from the set A, and includes 1,560 items.

All these patterns focus on three forms of *conjunctions*, namely '*and*', '*or*', and '*but not*', which connect two *conjuncts* (the extension of which varies in our data from one up to maximally eight tokens as in '*London's "Big Bang" 1986 deregulation and Toronto's "Little Bang" the same year*'.

The remainders from the conjunctions and the two conjuncts in a coordinated NP are called *shared elements* (e.g., '*owner*' and '*a happy*' in the above example). It is evident that the correct recognition of conjunct boundaries allows for the proper identification of the shared elements.

Set A contains 1,455 coordinations where N_1 and N_3 are coordinated (e.g, '*food and household goods*') and 1,232 coordinations where N_1 and N_2 are coordinated (e.g., '*cotton and acetate fibers*'). Set B consists of 643 coordinations where N_1 and N_3 are coordinated and 917 coordinations where N_1 and N_2 are coordinated.

The extracted data sets were converted into an IO representation of tokens labeled as '*C*' for conjunct, '*CC*' for conjunction, and '*S*' for the shared element(s). The noun phrase '*cotton and acetate fibers*', e.g., is represented as a sequence '*C CC C S*', while '*food and household goods*' is represented as a sequence '*C CC C C*'.

3 Methods

We here compare three different approaches to the resolution of noun phrase coordination ambiguity, *viz.* ones relying solely on morpho-syntactic information, solely on lexico-semantic information, and a cumulative combination of both. As far as semantic information is concerned we make use of various WORDNET similarity measures.

3.1 Baselines

We used three baselines for resolving noun phrase coordination ambiguities — one incorporating only lexico-semantic information, the *WordNet Similarity* baseline, and two alternative ones incorporating only morpho-syntactic and syntactic parse information, the *Number Agreement* and the *Bikel Parser* baseline, respectively.

3.1.1 WORDNET Similarity (WN) Baseline

Our lexico-semantic baseline comes with WORDNET semantic similarity scores of putatively coordinated nouns. For our experiments, we used the implementation of WORDNET similarity and relatedness measures provided by Ted Pedersen.[2] The following similarity measures were considered: two measures based on path lenghts between concepts (*path* and *lch* (Leacock et al., 1998)), three measures based on information content, i.e., corpus-based measures of the specificity of a concept (*res* (Resnik, 1999), *lin* (Lin, 1998), and *jcn* (Jiang and Conrath, 1997)). Furthermore, we used two relatedness measures, namely, *lesk* (Banerjee and Pedersen, 2003) and *vector* (Patwardhan et al., 2003), which score the similarity of the glosses of both concepts. We applied these similarity measures to any pair of putatively coordinated nouns in the noun phrases from our data sets, *A* and *B*. To determine potential conjuncts we calculate two similarity scores relative to the structures discussed in Section 2.2:

$$s_1 = sim(N_1, N_2) \text{ and } s_2 = sim(N_1, N_3)$$

Our final score is the maximum over both scores which is then the semantic indicator for the most plausible resolution of the coordination.

3.1.2 Number Agreement (NA) Baseline

We compared here the number agreement between selected nouns (see Resnik (1999)). Accordingly, N_1 and N_2 are coordinated, if $number(N_1) = number(N_2)$ AND $number(N_1) \neq number(N_3)$, while N_1 and N_3 are coordinated, if $number(N_1) = number(N_3)$ AND $number(N_1) \neq number(N_2)$.

3.1.3 Post-Processing Heuristics

In the *WN* and *NA* baselines, after the detection of coordinated elements we used simple heuristics to tag the remaining part of the noun phrase. If N_1 and N_2 were hypothesized to be coordinated, then all tokens preceding N_1 were tagged as

shared elements, N_3 was tagged as shared element as well, while all tokens between the conjunction and N_2 were tagged as conjuncts. For example, in '*a happy dog*$_{N_1}$ *and cat*$_{N_2}$ *owner*$_{N_3}$' we identify '*dog*' and '*cat*' as coordinated elements and tag '*a happy*' and '*owner*' as shared elements. The final resolution looks like '*S S C CC C S*'. If N_1 and N_3 were hypothesized to be coordinated, then all other elements except conjunctions were tagged as parts of conjuncts, as well.

3.1.4 Bikel Parser (BP) Baseline

We used the well-known Bikel Parser (Bikel, 2004) in its original version and the one used by Collins (2003). We trained both of them only with NPs extracted from the re-annotated version of WSJ (see Section 2) and converted the bracketing output of the parsers to the IO representation for NP coordinations for further evaluations.

3.2 Chunking of Conjuncts with CRFs

The approach to conjunct identification presented by Buyko et al. (2007) employs Conditional Random Fields (CRF) (Lafferty et al., 2001),[3] which assign a label to each token of coordinated NPs according to its function in the coordination: '*C*' for conjuncts, '*CC*' for conjunctions, and '*S*' for shared elements. Since non-nested conjuncts can be assumed to be in a sequential order, sequential learning approaches (instead of single position classification approaches) seem appropriate here.

Buyko et al. (2007) report an F-measure of 93% on conjunct identification in the GENIA corpus. They use a feature set including lexical (words), and morpho-syntactic features (POS tags, morpho-syntactic similarity of putative conjuncts), but exclude any semantic criteria. The morpho-syntactic similarity features were generated from a rule-based approach to conjunct identification using the maximal symmetry of conjuncts as constituted by their respective POS annotation.

We here intend to apply this approach for resolving coordination ambiguities involving noun compounds in the newswire language such as '*president and chief executive*'. This restricts the spectrum of considered coordinations in noun phrases to more complicated cases than those considered by Buyko et al. (2007). We will thus test the various resolution models under harder test conditions,

---

Feature Class	Description
default feature	prior probability distribution over the possible argument labels
lexical	word
morpho-syntactic	the token's POS tag; output labels of the morpho-syntactic similarity ('C','CC' and 'S') (see Buyko et al. (2007)); output labels of the number agreement baseline ('C', 'CC' and 'S')
semantic WN	output labels of the WORDNET similarity baseline ('C', 'CC' and 'S')
contextual	conjunctions of all features of neighboring tokens (two tokens to the left and one token to the right)

Table 1: Feature Classes Used for Conjunct Identification

since Buyko et al. consider, e.g., adjective coordinations in noun phrases such as *'positive and negative IL-2 regulator'* that are predominant in the biomedical language domain.

We also propose in this work an extension of the feature space in terms of lexico-semantic features (see Table 1), information that originates from similarity computations on WORDNET data. Furthermore, we do not use orthographical features of the original approach as they are well suited only for the biomedical language domain.

4 Results and Error Analysis

To evaluate the different approaches to conjunct identification, we used recall and precision scores since they are well suited for the evaluation of segmentation tasks. Two types of decisions were evaluated – the assignment of *'C'* labels denoting conjuncts in terms of the F-measure, and (given the tagged conjuncts) the *accuracy* of the complete coordination resolution. A coordination is resolved properly only, if all tokens of both conjuncts are correctly identified.

We carried out a ten-fold cross-validation of all ML-based methods (Bikel parser (Bikel, 2004) and CRF-based conjunct identification (Buyko et al., 2007)). For the evaluation of the *NA* and *WN* baselines, we tested their performance on the complete data sets, *A* and *B* (see Section 2).

As Table 2 depicts the *NA* baseline achieved an accuracy of 28.4% on *A* (36.6% on *B*), the Bikel parser reached 77.2% on *A* (73.4% in *B*), while the *WN* baseline got in its best run (*vector* measure) an accuracy of 41.7% on *A* (49.6% on *B*). These results already reveal that parsing almost dramatically outperforms the coordination resolution based on the *NA* similarity by up to 35.5%

points. The results of the *WN* baseline indicate that the best similarity measure for conjunct identification is the *vector* similarity (Patwardhan et al., 2003) that scores the similarity between the glosses of the concepts.

Our error analysis of the *WN* baseline on the test set *A* reveals that its low accuracy has various reasons. First, about 37% of the coordinations could not be resolved due to the absence of at least one noun involved in the coordination from the WORDNET. These coordinations usually include named entities such as person and organization names (e.g., *'brothers Francis and Gilbert Gros'*). These coverage gaps have clearly a negative effect on the resolution results for all WORDNET similarity measures.

To find out errors which are specific for the considered similarity measures, we have chosen the *res* measure and inspected the analysis results on all noun phrases where nouns are covered by WORDNET. The remaining set of coordinations contains 1,740 noun phrases. 1,022 coordinations (59%) of this set were completely resolved by the *WN* baseline, while 1,117 coordinations (64% of the remaining part, 41.5% of the test set *A*) could be at least partly resolved. Obviously, the coordinated heads are properly detected by the *res* measure but our heuristics for tagging the remaining modifiers (see Subsection 3.1.3) fail to provide the correct conjunct boundaries.

623 coordinations (36%) were mis-classified by the *res* measure. A closer look at this data reveals two types of errors. The first and minor type is the misleading selection of putatively coordinated heads N_1, N_2, and N_3. We presuppose in the *WN* baseline that the heads appear rightmost in the noun phrase, although that is not always the case as illustrated by the phrase *'Northern California earthquake and Hurricane Hugo'*. The *res* measure detected correctly a higher similarity between *'earthquake'* (N_1) and *'hurricane'* (N_2), but *'Hugo'* (N_3) is a modifier of *'hurricane'*. Although the *res* measure works fine, the coordination cannot be properly resolved due to syntactic reasons. In some cases, N_2 is wrongly selected as in *'life and health insurance operation'* where the *WN* baseline selects *'insurance'* as right-most noun (except the last noun *'operation'*) and not *'health'*.[4]

---

[4] *'turbine$_{N_2}$'* is, however, correctly selected in *'steam turbine and gas turbine plants'*.

	Set A		Set B	
	Recall/Precision/F-Score	Accuracy	Recall/Precision/F-Score	Accuracy
NA	32.7 / 75.9 / 45.7	28.4	41.6 / 83.9 / 55.6	36.6
Bikel	85.6 / 85.4 / 85.5	77.2	83.8 / 83.6 / 83.7	73.4
Bikel (Collins)	85.9 / 85.7 / 85.8	77.5	83.6 83.4 / 83.5	72.9
WN jcn	45.6 / 69.3 / 55.0	36.2	54.7 / 72.1 / 62.2	41.2
WN lch	48.7 / 70.8 / 57.7	39.2	57.8 / 74.1 / 65.0	44.4
WN lesk	49.2 / 66.2 / 56.5	38.1	59.3 / 70.3 / 64.3	43.3
WN lin	44.7 / 69.9 / 54.6	35.5	53.5 / 72.6 / 61.6	40.5
WN res	45.9 / 71.8 / 56.0	37.4	55.7 / 75.8 / 64.2	43.8
WN path	48.7 / 70.8 / 57.7	39.2	57.8 / 74.1 / 65.0	44.4
WN vector	51.2 / 68.9 / 58.8	**41.7**	62.8 / 74.5 / 68.1	**49.6**
CRF (default), contextual	75.2 / 72.4 / 73.8	60.3	77.9 / 75.1 / 76.5	63.1
+ Lexical, morpho-syntactic	87.1 / 87.2 / 87.1	**77.9**	88.1 / 88.0 / 88.0	**78.8**
+ WN (lesk)	87.2 / 87.2 / 87.2	**78.0**	88.2 / 88.2 / 88.2	**79.1**
CRF (default), contextual + only WN (lesk)	79.3 / 78.4 / 78.9	64.8	81.2 / 80.6 / 80.9	66.9
CRF (default), contextual + only morpho-syntactic	86.2 / 86.3 / 86.3	76.6	87.6 / 87.6 / 87.6	78.1

Table 2: F-measure of Conjunct Identification and Accuracy of Coordination Resolution on the WSJ Section of the PENN TREEBANK Corpus

The second type of error comes as erroneous classifications of the *res* measure such as in *'hospitals and blood banks'* where *'hospitals'* and *'blood'* have a higher similarity than *'hospitals'* and *'banks'* although they are, in fact, coordinated here. *'hotels and large restaurant chains'*, *'records and music publishing'*, *'chemicals and textiles company'* are other examples for the observation that the coordinated elements have a lower similarity as non-coordinated ones.

We also carried out a ten-fold cross-validation of the CRF-based approach for the conjunct identification. First of all, the CRF-based approach (with and without *WN* similarity) achieved the highest accuracy score – up to 78.0% on set *A*, and 79.1% on *B* – compared with all other approaches we scrutinized on. We also tested the performance of the original semantics-free approach and the additional effects of the WORDNET similarity measures (see Table 2). Although the integration of semantic information leads to a mild gain compared with the original approach (up to 0.3% points, with the *lesk* measure), the results indicate that no substantial benefit can be traced to semantic features.

We ran several tests with solely morpho-syntactic features (as enumerated in Table 1) and solely *WN* features, too. They reveal that solely morpho-syntactic features are up to 11.8% points more predictive than *WN* features. The best results were still achieved using the gloss-oriented *lesk* measure (see Table 3).

The inspection of the errors types from the various runs is not fully conclusive though. After adding *WN* features to both sets, we detected some improvements for conjunct tagging with high *WN* similarity. Some conjunct boundaries could be corrected as in *'record and movie producer'* where, in the first run, *'producer'* was tagged as a conjunct and was corrected as being shared by integrating *WN* features. But we also detected a degrading tagging behavior of conjuncts with *WN* features where the *WN* similarity was not helpful at all as in *'chairman and chief designer'* where *'chairman'* and *'chief'* under the influence of *WN* features were judged to be conjuncts. We found out that the addition of *WN* features positively influences the classification of coordinations where N_1 and N_2 are coordinated, while it increased errors in the classification of coordinations where N_1 and N_3 are coordinated.

In addition, we calculated intersections between the set *A* error data (unique) of the *res WN* baseline (1400 phrases) and the error data of the CRF approach without *WN* features (391), and the error data of the CRF approach with *WN* features (385), respectively. These error data sets contain noun phrases where coordinated heads could not be properly detected. The set of the *res WN* baseline and the set of the CRF approach without *WN* features have an intersection of 230 instances, where 138 instances could not be found in the WORDNET. That means that for about 161 instances (59%) in the mis-classified data of the CRF approach the additional *WN* features would not be helpful. The intersection remains similar (226 instances) between the set of the *res WN* base-

Default, Context, Lexical, Morpho-Syntactic Feats.	WN Sim	Recall Precision / F-Score	Accu-racy
∨		88.1/ 88.0 / 88.0	78.8
∨	jcn	87.9/ 87.8 / 87.9	78.3
∨	lch	87.9 / 87.9 / 87.9	78.5
∨	**lesk**	88.2/ 88.2 / 88.2	**79.1**
∨	lin	88.0 / 88.0 / 88.0	78.8
∨	res	88.2 / 88.2 / 88.2	79.0
∨	path	87.9 / 87.9 / 87.9	78.5
∨	vector	87.8 / 87.7 / 87.7	78.2

Table 3: Conjunct Identification — Cross-validation on the WSJ section of the PENN TREE-BANK Corpus on Test Set *B*

line and the set of the CRF approach enriched with *WN* features. The intersection between the errror sets of the both CRF approaches includes 352 instances. The integration of the *WN* features was not helful for almost the complete error data from the original CRF approach. We have previously shown that the *res WN* baseline features correlate with the correct label sequence for only 1,117 coordinations (41.5%) of the complete evaluation set *A* and the features thus do not seem to be effective in our approach.

Furthermore, we evaluated the results of the CRF approach only for the correct detection of coordinated heads (see above for the *res* measure and intersection counts) and disregarded the modifier classification. The results — 85.3% on set *A* and 85.4% on set *B* — reveal that the classification of modifiers is a major source of classification errors. In both configurations the problematic noun phrases are the ones with (e.g., adjectival) modifiers. The boundaries of conjuncts are not properly recognized in such noun phrases, as for example in *'American comic book and television series'* where the correct label sequence is *'S C C CC C S'*, since *'American'* is the shared modifier of *'book'* and *'television'*, while *'comic'* just modifies *'book'*.

As most adjectives appearing at the beginning of the noun phrase as in *'medical products and services company'* tend to be used as shared modifiers of coordinations in our data, this, erroneously, leads to false taggings, e.g., *'personal'* in *'personal computer and software design'* as a shared element. To cope adequately with modifiers we need to integrate more appropriate features such as collocation frequencies of modifiers and coordinated heads. The detection of a higher collocation frequency of *'personal computer'* in comparison to *'personal software'* (e.g., using the pro-

cedures proposed by Wermter and Hahn (2004)) would help tagging the conjunct boundaries.

5 Conclusions and Future Work

We investigated the problem of noun phrase coordination resolution as a segmentation problem among conjuncts involved in the coordination. While resolving coordination ellipsis is often considered as a semantically constrained problem, we wanted to assess a less 'costly' solution strategy, namely relying on 'cheaper' to get syntactic criteria as much as possible, though not sacrificing the accurary of resolutions.

We, first looked at morpho-syntactic criteria only and lexico-semantic criteria only, and then at the combination of both approaches. The evaluation results from a variety of experiments reveal that the major part of ambiguous coordinations can be resolved using solely morpho-syntactic features. Surprising as it might be, the semantic information as derived from the WORDNET similarity measures does not yield any further substantial improvement for our approach. This is somehow counter-intuitive, but our findings, unlike those from earlier studies which emphasized the role of semantic criteria, are based on extensive corpus data – the PENN TREEBANK.

Results from our error analysis will guide future work to further boost results. Particular emphasis will be laid on the integration of named entity recognizers, collocation frequencies and distributional similarity data as also advocated by Chantree et al. (2005).

The presented sequential labeling-based approach to coordination resolution was here applied to the resolution of a special type of ambiguous noun phrases. In general, this approach can easily be applied to the resolution of other types of coordinative structures in noun phrases as already presented in Buyko et al. (2007). As far as other phrasal types (e.g., verbal phrases) are concerned, long-distance coordinations play a much more prominent role. The token-based labeling approach may be thus substituted by a chunk-based approach operating on sentences.

References

Agarwal, R. and L. Boggess. 1992. A simple but useful approach to conjunct identification. In *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics*, pages 15–21. Newark, DE, USA, 28 June - 2 July 1992.

Banerjee, S. and T. Pedersen. 2003. Extended gloss overlaps as a measure of semantic relatedness. In *IJCAI'03 – Proceedings of the 18th International Joint Conference on Artificial Intelligence*, pages 805–810. Acapulco, Mexico, August 9-15, 2003.

Bikel, D. 2004. Intricacies of Collins' parsing model. *Computational Linguistics*, 30(4):479–511.

Buyko, E., K. Tomanek, and U. Hahn. 2007. Resolution of coordination ellipses in biological named entities using Conditional Random Fields. In *PACLING 2007 - Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics*, pages 163–171. Melbourne, Australia, September 19-21, 2007.

Chantree, F. A. Kilgarriff, A. de Roeck, and A. Willis. 2005. Disambiguating coordinations using word distribution information. In *RANLP 2005 – Proceedings of the Intl. Conference on 'Recent Advances in Natural Language Processing'*, pages 144–151. Borovets, Bulgaria, 21-23 September, 2005.

Charniak, E. and M. Johnson. 2005. Coarse-to-fine *n*-best parsing and MaxEnt discriminative reranking. In *ACL'05 – Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pages 173–180. Ann Arbor, MI, 25-30 June 2005.

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

Fellbaum, C., editor. 1998. WORDNET*: An Electronic Lexical Database*. MIT Press.

Hogan, D. 2007a. Coordinate noun phrase disambiguation in a generative parsing model. In *ACL'07 – Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 680–687. Prague, Czech Republic, June 28-29, 2007.

Hogan, D. 2007b. Empirical measurements of lexical similarity in noun phrase conjuncts. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics. Demo and Poster Sessions*, pages 149–152. Prague, Czech Republic, June 28-29, 2007.

Jiang, J. and D. Conrath. 1997. Semantic similarity based on corpus statistics and lexical taxonomy. In *ROCLING-X – Proceedings of the 1997 International Conference on Research in Computational Linguistics*. Taipei, Taiwan, August 22-24, 1997.

Lafferty, J., A. McCallum, and F. Pereira. 2001. Conditional Random Fields: Probabilistic models for segmenting and labeling sequence data. In *ICML-2001 – Proceedings of the 18th International Conference on Machine Learning*, pages 282–289. Williams College, MA, USA, June 28 - July 1, 2001.

Leacock, C., M. Chodorow, and G. Miller. 1998. Using corpus statistics and WORDNET relations for sense identification. *Computational Linguistics*, 24(1):147–165.

Lin, D. 1998. An information-theoretic definition of similarity. In *Proceedings of the 15th International Conference on Machine Learning*, pages 296–304. Madison, WI, USA, July 24-27, 1998.

Marcus, M., B. Santorini, and M.-A. Marcinkiewicz. 1993. Building a large annotated corpus of English: The PENN TREEBANK. *Computational Linguistics*, 19(2):313–330.

Nakov, P. and M. Hearst. 2005. Using the Web as an implicit training set: Application to structural ambiguity resolution. In *HLT-EMNLP'05 – Proceedings of the 5th Human Language Technology Conference and 2005 Conference on Empirical Methods in Natural Language Processing*, pages 835–842. Vancouver, B.C., Canada, October 6-8, 2005.

Ohta, T., Y. Tateisi, and J.-D. Kim. 2002. The GENIA corpus: An annotated research abstract corpus in molecular biology domain. In *HLT 2002 – Proceedings of the 2nd International Conference on Human Language Technology Research*, pages 82–86. San Diego, CA, USA, March 24-27, 2002.

Okumura, A. and K. Muraki. 1994. Symmetric pattern matching analysis for English coordinate structures. In *ANLP 1994 – Proceedings of the 4th Conference on Applied Natural Language Processing*, pages 41–46. Stuttgart, Germany, 13-15 October 1994.

Patwardhan, S., S. Banerjee, and T. Pedersen. 2003. Using measures of semantic relatedness for word sense disambiguation. In *CICLing 2003 – Proceedings 4th Intl. Conference on Computational Linguistics and Intelligent Text Processing*, pages 241–257. Mexico City, Mexico, February 16-22, 2003.

Resnik, P. 1999. Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *Journal of Artificial Intelligence Research*, 11:95–130.

Rus, V., D. Moldovan, and O. Bolohan. 2002. Bracketing compound nouns for logic form derivation. In *FLAIRS 2002 – Proceedings of the 15th International Florida Artificial Intelligence Research Society Conference*, pages 198–202. Pensacola Beach, FL, USA, May 14-16, 2002.

Shimbo, M. and K. Hara. 2007. A discriminative learning model for coordinate conjunctions. In *EMNLP-CoNLL 2007 – Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 610–619. Prague,Czech Republic, June 28-29, 2007.

Vadas, D. and J. Curran. 2007. Adding noun phrase structure to the PENN TREEBANK. In *ACL'07 – Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 240–247. Prague, Czech Republic, June 28-29, 2007.

Wermter, J. and U. Hahn. 2004. Collocation extraction based on modifiability statistics. In *COLING 2004 – Proceedings of the 20th International Conference on Computational Linguistics*, pages 980–986. Geneva, Switzerland, August 23-27, 2004.