Coordinate Noun Phrase Disambiguation in a Generative Parsing Model

Deirdre Hogan*
Computer Science Department
Trinity College Dublin
Dublin 2, Ireland
dhogan@computing.dcu.ie

Abstract

In this paper we present methods for improving the disambiguation of noun phrase (NP) coordination within the framework of a lexicalised history-based parsing model. As well as reducing noise in the data, we look at modelling two main sources of information for disambiguation: symmetry in conjunct structure, and the dependency between conjunct lexical heads. Our changes to the baseline model result in an increase in NP coordination dependency *f*-score from 69.9% to 73.8%, which represents a relative reduction in *f*-score error of 13%.

1 Introduction

Coordination disambiguation is a relatively little studied area, yet the correct bracketing of coordination constructions is one of the most difficult problems for natural language parsers. In the Collins parser (Collins, 1999), for example, dependencies involving coordination achieve an *f*-score as low as 61.8%, by far the worst performance of all dependency types.

Take the phrase *busloads of executives and their wives* (taken from the WSJ treebank). The coordinating conjunction (CC) *and* and the noun phrase *their wives* could attach to the noun phrase *executives*, as illustrated in Tree 1, Figure 1. Alternatively, *their wives* could be incorrectly conjoined to the noun phrase *busloads of executives* as in Tree 2, Figure 1.

As with PP attachment, most previous attempts at tackling coordination as a subproblem of parsing have treated it as a separate task to parsing and it is not always obvious how to integrate the methods proposed for disambiguation into existing parsing models. We therefore approach coordination disambiguation, not as a separate task, but from within the framework of a generative parsing model.

As noun phrase coordination accounts for over 50% of coordination dependency error in our baseline model we focus on NP coordination. Using a model based on the generative parsing model of (Collins, 1999) Model 1, we attempt to improve the ability of the parsing model to make the correct coordination decisions. This is done in the context of parse reranking, where the *n*-best parses output from Bikel's parser (Bikel, 2004) are reranked according to a generative history-based model.

In Section 2 we summarise previous work on coordination disambiguation. There is often a considerable bias toward symmetry in the syntactic structure of two conjuncts and in Section 3 we introduce new parameter classes to allow the model to prefer symmetry in conjunct structure. Section 4 is concerned with modelling the dependency between conjunct head words and begins by looking at how the different handling of coordination in noun phrases and base noun phrases (NPB) affects coordination disambiguation.[1] We look at how we might improve the model's handling of coordinate head-head dependencies by altering the model so that a common

* Now at the National Centre for Language Technology, Dublin City University, Ireland.

[1] A base noun phrase, as defined in (Collins, 1999), is a noun phrase which does not directly dominate another noun phrase, unless that noun phrase is possessive.

Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, pages 680–687,
Prague, Czech Republic, June 2007.

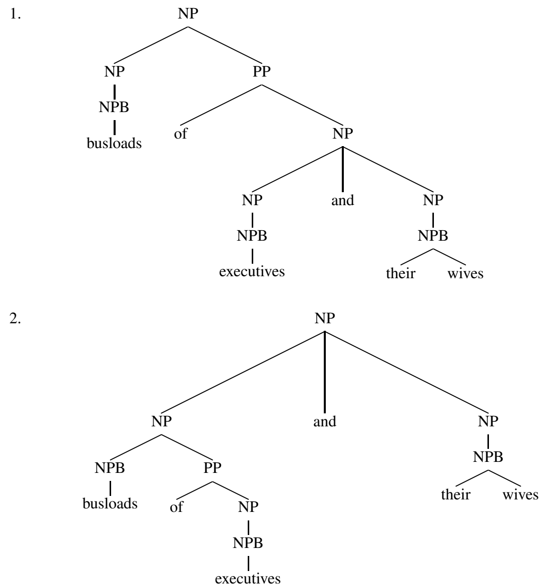

Figure 1: Tree 1. The correct noun phrase parse. Tree 2. The incorrect parse for the noun phrase.

parameter class is used for coordinate word probability estimation in both NPs and NPBs. In Section 4.2 we focus on improving the estimation of this parameter class by incorporating BNC data, and a measure of word similarity based on vector cosine similarity, to reduce data sparseness. In Section 5 we suggest a new head-finding rule for NPBs so that the lexicalisation process for coordinate NPBs is more similar to that of other NPs.

Section 6 examines inconsistencies in the annotation of coordinate NPs in the Penn Treebank which can lead to errors in coordination disambiguation. We show how some coordinate noun phrase inconsistencies can be automatically detected and cleaned from the data sets. Section 7 details how the model is evaluated, presents the experiments made and gives a breakdown of results.

2 Previous Work

Most previous attempts at tackling coordination have focused on a particular type of NP coordination to disambiguate. Both Resnik (1999) and Nakov and Hearst (2005) consider NP coordinations of the form *n1 and n2 n3* where two structural analyses are possible: *((n1 and n2) n3)* and *((n1) and (n2 n3))*. They aim to show more structure than is shown in trees following the Penn guidelines, whereas in our approach we aim to reproduce Penn guideline trees. To resolve the ambiguities, Resnik combines number agreement information of candidate conjoined nouns, an information theoretic measure of semantic similarity, and a measure of the appropriateness of noun-noun modification. Nakov and Hearst (2005) disambiguate by combining Web-based statistics on head word co-occurrences with other mainly heuristic information sources.

A probabilistic approach is presented in (Goldberg, 1999), where an unsupervised maximum entropy statistical model is used to disambiguate coordinate noun phrases of the form *n1 preposition n2 cc n3*. Here the problem is framed as an attachment decision: does *n3* attach 'high' to the first noun, *n1*, or 'low' to *n2*?

In (Agarwal and Boggess, 1992) the task is to identify pre-CC conjuncts which appear in text that has been part-of-speech (POS) tagged and semi-parsed, as well as tagged with semantic labels specific to the domain. The identification of the pre-CC conjunct is based on heuristics which choose the pre-CC conjunct that maximises the symmetry between pre- and post-CC conjuncts.

Insofar as we do not separate coordination disambiguation from the overall parsing task, our approach resembles the efforts to improve coordination disambiguation in (Kurohashi, 1994; Ratnaparkhi, 1994; Charniak and Johnson, 2005). In (Kurohashi, 1994) coordination disambiguation is carried out as the first component of a Japanese dependency parser using a technique which calculates similarity between series of words from the left and right of a conjunction. Similarity is measured based on matching POS tags, matching words and a thesaurus-based measure of semantic similarity. In both the discriminative reranker of Ratnaparkhi et al. (1994) and that of Charniak and Johnson (2005) features are included to capture syntactic parallelism across conjuncts at various depths.

3 Modelling Symmetry Between Conjuncts

There is often a considerable bias toward symmetry in the syntactic structure of two conjuncts, see for example (Dubey et al., 2005). Take Figure 2: If we take as level 0 the level in the coordinate sub-

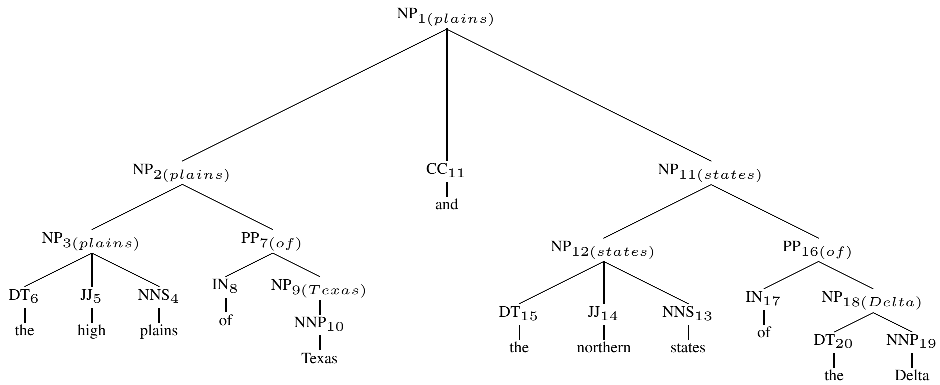

Figure 2: Example of symmetry in conjunct structure in a lexicalised subtree.

tree where the coordinating conjunction *CC* occurs, then there is exact symmetry in the two conjuncts in terms of non-terminal labels and head word part-of-speech tags for levels 0, 1 and 2. Learning a bias toward parallelism in conjuncts should improve the parsing model's ability to correctly attach a coordination conjunction and second conjunct to the correct position in the tree.

In history-based models, features are limited to being functions of the tree generated so far. The task is to incorporate a feature into the model that captures a particular bias yet still adheres to derivation-based restrictions. Parses are generated top-down, head-first, left-to-right. Each node in the tree in Figure 2 is annotated with the order the nodes are generated (we omit, for the sake of clarity, the generation of the *STOP* nodes). Note that when the decision to attach the second conjunct to the head conjunct is being made (i.e. Step 11, when the *CC* and *NP(states)* nodes are being generated) the subtree rooted at *NP(states)* has not yet been generated. Thus at the point that the conjunct attachment decision is made it is not possible to use information about symmetry of conjunct structure, as the structure of the second conjunct is not yet known.

It is possible, however, to condition on structure of the already generated head conjunct when building the internal structure of the second conjunct. In our model when the structure of the second conjunct is being generated we condition on features which are functions of the first conjunct. When generating a node N_i in the second conjunct, we retrieve the corresponding node $N_{i_{preCC}}$ in the first conjunct, via a left to right traversal of the first conjunct. For example, from Figure 2 the pre-*CC* node *NP(Texas)* is the node corresponding to *NP(Delta)* in the post-*CC* conjunct. From $N_{i_{preCC}}$ we extract information, such as its part-of-speech, for use as a feature when predicting a POS tag for the corresponding node in the post-*CC* conjunct.

When generating a second conjunct, instead of the usual parameter classes for estimating the probability of the head label C_h and the POS label of a dependent node t_i, we created two new parameter classes which are used only in the generation of second conjunct nodes:

$$P_{ccC_h}(C_h|\gamma(headC), C_p, w_p, t_p, t_{gp}, depth) \tag{1}$$

$$P_{cct_i}(t_i|\alpha(headC), dir, C_p, w_p, t_p, dist, t_{i\_1}, t_{i\_2}, depth) \tag{2}$$

where $\gamma(headC)$ returns the non-terminal label of $N_{i_{preCC}}$ for the node in question and $\alpha(headC)$ returns the POS tag of $N_{i_{preCC}}$. Both functions return *+NOMATCH+* if there is no $N_{i_{preCC}}$ for the node. Depth is the level of the post-*CC* conjunct node being generated.

4 Modelling Coordinate Head Words

Some noun pairs are more likely to be conjoined than others. Take again the trees in Figure 1. The two head nouns coordinated in Tree 1 are *executives* and *wives*, and in Tree 2: *busloads* and *wives*. Clearly, the former pair of head nouns is more likely and, for the purpose of discrimination, the model would benefit if it could learn that *executives and wives* is a more likely combination than *busloads and wives*.

Bilexical head-head dependencies of the type found in coordinate structures are a somewhat different class of dependency to modifier-head dependencies. In *the fat cat*, for example, there is clearly one head to the noun phrase: *cat*. In *cats and dogs* however there are two heads, though in the parsing model just one is chosen, somewhat arbitrarily, to head the entire noun phrase.

In the baseline model there is essentially one parameter class for the estimation of word probabilities:

$$P_{word}(w_i|H(i)) \tag{3}$$

where w_i is the lexical head of constituent i and $H(i)$ is the history of the constituent. The history is made up of conditioning features chosen from structure that has already been determined in the top-down derivation of the tree.

In Section 4.1 we discuss how though the coordinate head-head dependency is captured for NPs, it is not captured for NPBs. We look at how we might improve the model's handling of coordinate head-head dependencies by altering the model so that a common parameter class in (4) is used for coordinate word probability estimation in both NPs and NPBs.

$$P_{coordWord}(w_i|w_p, H(i)) \tag{4}$$

In Section 4.2 we focus on improving the estimation of this parameter class by reducing data sparseness.

4.1 Extending $P_{coordWord}$ to Coordinate NPBs

In the baseline model each node in the tree is annotated with a coordination flag which is set to true for the node immediately following the coordinating conjunction. For coordinate NPs the head-head dependency is captured when this flag is set to true. In Figure 1, discarding for simplicity the other features in the history, the probability of the coordinate head *wives*, is estimated in Tree 1 as:

$$P_{word}(w_i = wives|coord = true, w_p = executives, ...) \tag{5}$$

and in Tree 2:

$$P_{word}(w_i = wives|coord = true, w_p = busloads, ...) \tag{6}$$

where w_p is the head word of the node to which the node headed by w_i is attaching and $coord$ is the coordination flag.

Unlike NPs, in NPBs (i.e. flat, non-recursive NPs) the coordination flag is not used to mark whether a node is a coordinated head or not. This flag is always set to false for NPBs. In addition, modifiers within NPBs are conditioned on the previously generated modifier rather than the head of the phrase.[2] This means that in an NPB such as *(cats and dogs)*, the estimate for the word *cats* will look like:

$$P_{word}(w_i = cats|coord = false, w_p = and, ...) \tag{7}$$

In our new model, for NPs, when the coordination flag is set to true, we use the parameter class in (4) to estimate the probability of one lexical head noun, given another. For NPBs, if a noun is generated directly after a CC then it is taken to be a coordinate head, w_i, and conditioned on the noun generated before the coordinating conjunction, which is chosen as w_p, and also estimated using (4).

4.2 Estimating the $P_{coordWord}$ parameter class

Data for bilexical statistics are particularly sparse. In order to decrease the sparseness of the coordinate head noun data, we extracted from the BNC examples of coordinate head noun pairs. We extracted all noun pairs occurring in a pattern of the form: *noun cc noun*, as well as lists of any number of nouns separated by commas and ending in *cc noun*.[3] To this data we added all head noun pairs from the WSJ that occurred together in a coordinate noun phrase, identified when the coordination flag was set to true. Every occurrence n_i *CC* n_j was also counted as an occurrence of n_j *CC* n_i. This further helps reduce sparseness.

The probability of one noun, n_i being coordinated with another n_j can be calculated simply as:

$$P_{lex}(n_i|n_j) = \frac{|n_i n_j|}{|n_j|} \tag{8}$$

Again to reduce data sparseness, we introduce a measure of word similarity. A word can be represented as a vector where every dimension of the vector represents another word type. The values of the vector components, the term weights, are derived from word co-occurrence counts. Cosine similarity between two word vectors can then be used to measure the similarity of two words. Measures of

[2] A full explanation of the handling of coordination in the model is given in (Bikel, 2004).

[3] Extracting coordinate noun pairs from the BNC in such a fashion follows work on networks of concepts described in (Widdows, 2004).

similarity between words based on similarity of co-occurrence vectors have been used before, for example, for word sense disambiguation (Schütze, 1998) and for PP-attachment disambiguation (Zhao and Lin, 2004). Our measure resembles that of (Caraballo, 99) where co-occurrence is also defined with respect to coordination patterns, although the experimental details in terms of data collection and vector term weights differ.

We can now incorporate the similarity measure into the probability estimate of (8) to give a new k-NN style method of estimating bilexical statistics based on weighting events according to the word similarity measure:

$$P_{sim}(n_i|n_j) = \frac{\sum_{n_x \in N(n_j)} sim(n_j, n_x)|n_i n_x|}{\sum_{n_x \in N(n_j)} sim(n_j, n_x)|n_x|} \quad (9)$$

where $sim(n_j, n_x)$ is a similarity score between words n_j and n_x and $N(n_j)$ is the set of words in the neighbourhood of n_j. This neighbourhood can be based on the k-nearest neighbours of n_j, where nearness is measured with the similarity function.

In order to smooth the bilexical estimate in (9) we combine it with another estimate, trained from WSJ data, by way of linear interpolation:

$$P_{coordWord}(n_i|n_j) = \lambda_{n_j} P_{sim}(n_i|n_j) + (1 - \lambda_{n_j}) P_{MLE}(n_i|t_i) \quad (10)$$

where t_i is the POS tag of word n_i, $P_{MLE}(n_i|t_i)$ is the maximum-likelihood estimate calculated from annotated WSJ data, and λ_{n_j} is calculated as in (11). In (11) we adapt the Witten-Bell method for the calculation of the weight λ, as used in the Collins parser, so that it incorporates the similarity measure for all words in the neighbourhood of n_j.

$$\lambda_{n_j} = \frac{\sum_{n_x \in N(n_j)} sim(n_j, n_x)|n_x|}{\sum_{n_x \in N(n_j)} sim(n_j, n_x)(|n_x| + CD(n_x))} \quad (11)$$

where C is a constant that can be optimised using held-out data and $D(n_j)$ is the diversity of a word n_j: the number of distinct words with which n_j has been coordinated in the training set.

The estimate in (9) can be viewed as the estimate with the more general history context than that of (8) because the context includes not only n_j but also words similar to n_j. The final probability estimate for $P_{coordWord}$ is calculated as the most specific estimate, P_{lex}, combined via regular Witten-Bell interpolation with the estimate in (10).

5 NPB Head-Finding Rules

Head-finding rules for coordinate NPBs differ from coordinate NPs.[4] Take the following two versions of the noun phrase *hard work and harmony*: (c) *(NP (NPB hard work and harmony))* and (d) *(NP (NP (NPB hard work)) and (NP (NPB harmony)))*. In the first example, *harmony* is chosen as head word of the NP; in example (d) the head of the entire NP is *work*. The choice of head affects the various dependencies in the model. However, in the case of two coordinate NPs which, as in the above example, cover the same span of words and differ only in whether the coordinate noun phrase is flat as in (c) or structured as in (d), the choice of head for the phrase is not particularly informative. In both cases the head words being coordinated are the same and either word could plausibly head the phrase; discrimination between trees in such cases should not be influenced by the choice of head, but rather by other, salient features that distinguish the trees.[5]

We would like to alter the head-finding rules for coordinate NPBs so that, in cases like those above, the word chosen to head the entire coordinate noun phrase would be the same for both base and non-base noun phrases. We experiment with slightly modified head-finding rules for coordinate NPBs. In an NPB such as *NPB → n1 CC n2 n3*, the head rules remain unchanged and the head of the phrase is (usually) the rightmost noun in the phrase. Thus, when *n2* is immediately followed by another noun the default is to assume nominal modifier coordination and the head rules stay the same. The modification we make to the head rules for NPBs is as follows: when *n2* is *not* immediately followed by a noun then the noun chosen to head the entire phrase is *n1*.

6 Inconsistencies in WSJ Coordinate NP Annotation

An inspection of NP coordination error in the baseline model revealed inconsistencies in WSJ annota-

[4] See (Collins, 1999) for the rules used in the baseline model.

[5] For example, it would be better if discrimination was largely based on whether *hard* modifies both *work* and *harmony* (c), or whether it modifies *work* alone (d).

tion. In this section we outline some types of coordinate NP inconsistency and outline a method for detecting some of these inconsistencies, which we later use to automatically clean noise from the data. Eliminating noise from treebanks has been previously used successfully to increase overall parser accuracy (Dickinson and Meurers, 2005).

The annotation of NPs in the Penn Treebank (Bies et al., 1995) follows somewhat different guidelines to that of other syntactic categories. Because their interpretation is so ambiguous, no internal structure is shown for nominal modifiers. For NPs with more than one head noun, if the only unshared modifiers in the constituent are nominal modifiers, then a flat structure is also given. Thus in *(NP the Manhattan phone book and tour guide)*[6] a flat structure is given because although *the* is a non-nominal modifier, it is shared, modifying both *tour guide* and *phone book*, and all other modifiers in the phrase are nominal.

However, we found that out of 1,417 examples of NP coordination in sections 02 to 21, involving phrases containing only nouns (common nouns or a mixture of common and proper nouns) and the coordinating conjunction, as many as 21.3%, contrary to the guidelines, were given internal structure, instead of a flat annotation. When all proper nouns are involved this phenomenon is even more common.[7]

Another common source of inconsistency in coordinate noun phrase bracketing occurs when a non-nominal modifier appears in the coordinate noun phrase. As previously discussed, according to the guidelines the modifier is annotated flat if it is shared. When the non-nominal modifier is unshared, more internal structure is shown, as in: *(NP (NP (NNS fangs)) (CC and) (NP (JJ pointed) (NNS ears)))*. However, the following two structured phrases, for example, were given a completely flat structure in the treebank: (a) *(NP (NP (NN oversight))(CC and) (NP (JJ disciplinary)(NNS procedures)))*, (b) *(NP (ADJP (JJ moderate)(CC and)(JJ low-cost))(NN housing))*. If we follow the guidelines then any coordinate NPB which ends with the following tag sequence can be automatically detected as incorrectly bracketed: *CC/non-nominal modifier/noun*. This is because either the non-nominal modifier, which is unambiguously unshared, is part of a noun phrase as (a) above, or it conjoined with another modifier as in (b). We found 202 examples of this in the training set, out of a total of 4,895 coordinate base noun phrases.

Finally, inconsistencies in POS tagging can also lead to problems with coordination. Take the bigram *executive officer*. We found 151 examples in the training set of a base noun phrase which ended with this bigram. 48% of the cases were POS tagged *JJ NN*, 52% tagged *NN NN*. [8] This has repercussions for coordinate noun phrase structure, as the presence of an adjectival pre-modifier indicates a structured annotation should be given.

These inconsistencies pose problems both for training and testing. With a relatively large amount of noise in the training set the model learns to give structures, which should be very unlikely, too high a probability. In testing, given inconsistencies in the gold standard trees, it becomes more difficult to judge how well the model is doing. Although it would be difficult to automatically detect the POS tagging errors, the other inconsistencies outlined above can be detected automatically by simple pattern matching. Automatically eliminating such examples is a simple method of cleaning the data.

7 Experimental Evaluation

We use a parsing model similar to that described in (Hogan, 2005) which is based on (Collins, 1999) Model 1 and uses *k*-NN for parameter estimation. The *n*-best output from Bikel's parser (Bikel, 2004) is reranked according to this *k*-NN parsing model, which achieves an *f*-score of 89.4% on section 23. For the coordination experiments, sections 02 to 21 are used for training, section 23 for testing and the remaining sections for validation. Results are for sentences containing 40 words or less.

As outlined in Section 6, the treebank guidelines are somewhat ambiguous as to the appropriate bracketing for coordinate NPs which consist entirely of proper nouns. We therefore do not include, in the coordination test and validation sets, coordinate NPs where in the gold standard NP the leaf nodes consist entirely of proper nouns (or CCs or commas). In do-

[6] In this section we do not show the NPB levels.

[7] In the guidelines it is recognised however that proper names are frequently annotated with internal structure.

[8] According to the POS bracketing guidelines (Santorini, 1991) the correct sequence of POS tags should be *NN NN*.

ing so we hope to avoid a situation whereby the success of the model is measured in part by how well it can predict the often inconsistent bracketing decisions made for a particular portion of the treebank.

In addition, and for the same reasons, if a gold standard tree is inconsistent with the guidelines in either of the following two ways the tree is not used when calculating coordinate precision and recall of the model: the gold tree is a noun phrase which ends with the sequence *CC/non-nominal modifier/noun*; the gold tree is a structured coordinate noun phrase where each word in the noun phrase is a noun.[9] Call these inconsistencies type *a* and type *b* respectively. This left us with a coordination validation set consisting of 1064 coordinate noun phrases and a test set of 416 coordinate NPs from section 23.

A coordinate phrase was deemed correct if the parent constituent label, and the two conjunct node labels (at level 0) match those in the gold subtree and if, in addition, each of the conjunct head words are the same in both test and gold tree. This follows the definition of a coordinate dependency in (Collins, 1999). Based on these criteria, the baseline *f*-scores for test and validation set were 69.1% and 67.1% respectively. The coordination *f*-score for the oracle trees on section 23 is 83.56%. In other words: if an 'oracle' were to choose from each set of *n*-best trees the tree that maximised constituent precision and recall, then the resulting set of oracle trees would have a NP coordination dependency *f*-score of 83.56%. For the validation set the oracle trees coordination dependency *f*-score is 82.47%.

7.1 Experiments and Results

We first eliminated from the training set all coordinate noun phrase subtrees, of type *a* and type *b* described in Section 7. The effect of this on the validation set is outlined in Table 1, step 2.

For the new parameter class in (1) we found that the best results occurred when it was used only in conjuncts of depth 1 and 2, although the case base for this parameter class contained head events from all post-*CC* conjunct depths. Parameter class (2) was used for predicting POS tags at level 1 in right-of-head conjuncts, although again the sample contained events from all depths.

[9]Recall from §6 that for this latter case the noun phrase should be flat - an NPB - rather than a noun phrase with internal structure.

Model	*f*-score	significance
1. Baseline	67.1	
2. NoiseElimination	68.7	$\gg 1$
3. Symmetry	69.9	$> 2, \gg 1$
4. NPB head rule	70.6	NOT $> 3, > 2, \gg 1$
5. $P_{coordWord}$ WSJ	71.7	NOT $> 4, > 3, \gg 2$
6. BNC data	72.1	NOT $> 5, > 4, \gg 3$
7. $sim(w_i, w_p)$	72.4	NOT > 6, NOT $> 5, \gg 4$

Table 1: Results on the Validation Set. 1064 coordinate noun phrase dependencies. In the significance column $>$ means at level .05 and $\gg$ means at level .005, for McNemar's test of significance. Results are cumulative.

For the $P_{coordWord}$ parameter class we extracted 9961 coordinate noun pairs from the WSJ training set and 815,323 pairs from the BNC. As pairs are considered symmetric this resulted in a total of 1,650,568 coordinate noun events. The term weights for the word vectors were dampened co-occurrence counts, of the form: $1 + log(count)$. For the estimation of $P_{sim}(n_i|n_j)$ we found it too computationally expensive to calculate similarity measures between n_j and each word token collected. The best results were obtained when the neighbourhood of n_j was taken to be the k-nearest neighbours of n_j from among the set of word that had previously occurred in a coordination pattern with n_j, where *k* is 1000. Table 1 shows the effect of the $P_{coordWord}$ parameter class estimated from WSJ data only (step 5), with the addition of BNC data (step 6) and finally with the word similarity measure (step 7).

The result of these experiments, as well as that involving the change in the head-finding heuristics, outlined in Section 5, was an increase in coordinate noun phrase *f*-score from 69.9% to 73.8% on the test set. This represents a 13% relative reduction in coordinate *f*-score error over the baseline, and, using McNemar's test for significance, is significant at the 0.05 level ($p = 0.034$). The reranker *f*-score for all constituents (not excluding any coordinate NPs) for section 23 rose slightly from 89.4% to 89.6%, a small but significant increase in *f*-score.[10]

Finally, we report results on an unaltered coordination test set, that is, a test set from which no

[10]Significance was calculated using the software available at www.cis.upenn.edu/ dbikel/software.html.

noisy events were eliminated. The baseline coordination dependency *f*-score for all NP coordination dependencies (550 dependencies) from section 23 is 69.27%. This rises to 72.74% when all experiments described in Section 7 are applied, which is also a statistically significant increase ($p = 0.042$).

8 Conclusion and Future Work

This paper outlined a novel method for modelling symmetry in conjunct structure, for modelling the dependency between noun phrase conjunct head words and for incorporating a measure of word similarity in the estimation of a model parameter. We also demonstrated how simple pattern matching can be used to reduce noise in WSJ noun phrase coordination data. Combined, these techniques resulted in a statistically significant improvement in noun phrase coordination accuracy.

Coordination disambiguation necessitates information from a variety of sources. Another information source important to NP coordinate disambiguation is the dependency between non-nominal modifiers and nouns which cross CCs in NPBs. For example, modelling this type of dependency could help the model learn that the phrase *the cats and dogs* should be bracketed flat, whereas the phrase *the U.S. and Washington* should be given structure.

Acknowledgements We are grateful to the TCD Broad Curriculum Fellowship scheme and to the SFI Basic Research Grant 04/BR/CS370 for funding this research. Thanks to Pádraig Cunningham, Saturnino Luz, Jennifer Foster and Gerard Hogan for helpful discussions and feedback on this work.

References

Rajeev Agarwal and Lois Boggess. 1992. A Simple but Useful Approach to Conjunct Identification. In *Proceedings of the 30th ACL.*

Ann Bies, Mark Ferguson, Karen Katz and Robert MacIntyre. 1995. Bracketing Guidelines for Treebank II Style Penn Treebank Project. Technical Report. University of Pennsylvania.

Dan Bikel. 2004. *On The Parameter Space of Generative Lexicalized Statistical Parsing Models*. Ph.D. thesis, University of Pennsylvania.

Sharon Caraballo. 1999. Automatic construction of a hypernym-labeled noun hierarchy from text. In *Proceedings of the 37th ACL.*

Eugene Charniak and Mark Johnson. 2005. Coarse-to-fine *n*-best Parsing and MaxEnt Discriminative Reranking. In *Proceedings of the 43rd ACL.*

Michael Collins. 1999. *Head-Driven Statistical Models for Natural Language Parsing*. Ph.D. thesis, University of Pennsylvania.

Markus Dickinson and W. Detmar Meurers. 2005. Prune diseased branches to get healthy trees! How to find erroneous local trees in a treebank and why it matters. In *Proceedings of the Fourth Workshop on Treebanks and Linguistic Theories (TLT).*

Amit Dubey, Patrick Sturt and Frank Keller. 2005. Parallelism in Coordination as an Instance of Syntactic Priming: Evidence from Corpus-based Modeling. In *Proceedings of the HLT/EMNP-05*.

Miriam Goldberg. 1999. An Unsupervised Model for Statistically Determining Coordinate Phrase Attachment. In *Proceedings of the 27th ACL.*

Deirdre Hogan. 2005. *k*-NN for Local Probability Estimation in Generative Parsing Models. In *Proceedings of the IWPT-05.*

Sadao Kurohashi and Makoto Nagao. 1994. A Syntactic Analysis Method of Long Japanese Sentences Based on the Detection of Conjunctive Structures. In *Computational Linguistics*, 20(4).

Preslav Nakov and Marti Hearst. 2005. Using the Web as an Implicit Training Set: Application to Structural Ambiguity Resolution. In *Proceedings of the HLT/EMNLP-05.*

Adwait Ratnaparkhi, Salim Roukos and R. Todd Ward. 1994. A Maximum Entropy Model for Parsing. In *Proceedings of the International Conference on Spoken Language Processing*.

Philip Resnik. 1999. Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language. In *Journal of Artificial Intelligence Research*, 11:95-130, 1999.

Beatrice Santorini. 1991. Part-of-Speech Tagging Guidelines for the Penn Treebank Project. Technical Report. University of Pennsylvania.

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

Dominic Widdows. 2004. *Geometry and Meaning*. CSLI Publications, Stanford, USA.

Shaojun Zhao and Dekang Lin. 2004. A Nearest-Neighbor Method for Resolving PP-Attachment Ambiguity. In *Proceedings of the IJCNLP-04.*