

Generalised PP-Attachment Disambiguation using Corpus-based Linguistic Diagnostics

Paola Merlo

Linguistics Department

University of Geneva

2 rue de Candolle

1211 Geneva 4, Switzerland

merlo@lettres.unige.ch

Abstract

We propose a new formulation of the PP attachment problem as a 4-way classification which takes into account the argument or adjunct status of the PP. Based on linguistic diagnostics, we train a 4-way classifier that reaches an average accuracy of 73.9% (baseline 66.2%). Compared to a sequence of binary classifiers, the 4-way classifier reaches better performance and individuates a verb's arguments more accurately, thus improving the acquisition of a crucial piece of information for many NLP applications.

1 Motivation

Incorrect attachment of prepositional phrases often constitutes the main source of errors in current parsing systems. Correct attachment of PPs is necessary to construct a parse tree which will support the proper interpretation of constituents in the sentence. Consider the time-worn example

I saw the man with the telescope

It is important to determine if the PP *with the telescope* is to be attached as a sister to the noun *the man*, restricting its interpretation, or if it is to be attached to the verb, thereby indicating the instrument of the main action described by the sentence. Based on examples of this sort, recent approaches have formalised the problem of disambiguating PP attachments as a binary choice, distinguishing between attachment of a PP to a given verb or to the verb's direct object (Ratnaparkhi et al., 1994; Collins and Brooks, 1995).

This is, however, a simplification of the problem, which does not take the nature of the attachment into account. Precisely, it does not distinguish PP arguments from PP adjuncts. Consider the following example, which contains two PPs, both modifying the verb.

Put the block on the table in the morning

The first PP is a locative PP required by the subcategorisation frame of the verb *put*, while *in the morning* is an optional descriptor of the time at which the action was performed. Though both attached to the verb, the two PPs entertain different relationships with the verb – the first is an argument while the latter is an adjunct. Analogous examples could be built for attachments to the noun.

Is it important to model not just the site but also the nature of the attachment of a PP into the tree structure? We would like to claim that it is. Distinguishing arguments from adjuncts is key to identifying the semantic kernel of a sentence. Extracting the core meaning of a sentence or phrase, in turn, is necessary for automatic acquisition of important lexical knowledge, such as subcategorisation frames and argument structures, which is used in several NLP tasks and applications, such as parsing or machine translation (Srinivas and Joshi, 1999; Dorr, 1997). Moreover, from a quantitative point of view, arguments and adjuncts have different statistical properties, requiring different statistical techniques. For example, (Hindle and Rooth, 1993) clearly indicate that their lexical association technique performs much better for arguments than for adjuncts, whether the attachment is to the verb or to the noun.

Researchers have abstracted away from this distinction, because identifying arguments and adjuncts is a notoriously difficult task, taxing many

native speakers' intuitions and requiring complex world knowledge. The usual expectation has been that this discrimination is not amenable to a corpus-based treatment. In recent work, however, we succeed in distinguishing arguments from adjuncts using evidence extracted from a parsed corpus (Merlo and Leybold, 2001). Our method develops corpus-based statistical correlates for the diagnostics used in linguistics to decide whether a PP is an argument or an adjunct. A numerical vectorial representation of the notion of argumenthood is provided, which supports automatic classification, reaching 86% accuracy. In the current paper, we extend this work and propose a new formulation of the PP attachment problem. We treat PP attachment as a 4-way classification of PPs into noun argument PPs, noun adjunct PPs, verb argument PPs, and verb adjunct PPs.

We show that it is possible to build a classifier that solves this problem using corpus evidence, with good accuracy (74%). This classifier solves the classification problem directly, in one step. Interestingly, we show that a 4-way classifier reaches better accuracy than a two-step sequence of binary classifiers, which first solve the noun-verb attachment problem and then refine the attachment decision into argument or adjunct. This result indicates that the current formulation of the PP attachment problem cannot be considered a first step in the solution of the final 4-way discrimination task. Finally, we note that the improvement is especially due to a better recognition of verbs' arguments, thus providing more accurate information to many NLP tasks and applications.

Solving this novel 4-way classification task crucially relies on the ability to distinguish arguments from adjuncts using corpus counts.

2 A Novel Method to Distinguish Arguments from Adjuncts

Few attempts have been made to distinguish arguments from adjuncts automatically (Buchholz, 1999; Merlo and Leybold, 2001; Villavicencio, 2002; Aldezabal et al., 2002). The core difficulty in this enterprise is to define the notion of argument precisely. There is a consensus in linguistics that arguments and adjuncts are different both with respect to their function in the sentence and in

the way they themselves are interpreted (Jackendoff, 1977; Marantz, 1984; Pollard and Sag, 1987; Grimshaw, 1990). With respect to their function, an argument fills a role in the relation described by its associated head, while an adjunct predicates a separate property of its associate head or phrase. With respect to their interpretation, a complement is an argument if its interpretation depends exclusively on the head with which it is associated, while it is an adjunct if its interpretation remains relatively constant when associating with different heads (Grimshaw, 1990, 108). Restricting the discussion to PPs, these differences are illustrated in the following examples (PP-argument in bold), see also (Schütze, 1995, 100).

- a) Kim camps/jogs/meditates on Sunday
- b) Kim depended **on Sandy**

In example a) the PP *on Sunday* can be construed without any reference to the preceding part of the sentence, and it preserves its meaning even when combining with different heads. This is, however, not the case for b). Here, the PP can only be properly understood in connection with the rest of the sentence: Sandy is the person on whom someone depends.

These semantic distinctions surface in observable syntactic differences, giving rise to a set of linguistic diagnostics to determine whether a PP is an adjunct or an argument. We illustrate here those countable diagnostics that can be approximated statistically and estimated using corpus counts, thus combining linguistic insight with the robustness of corpus-based methods.

3 The Linguistic Diagnostics

Many diagnostics for argumenthood have been proposed in the literature (Schütze, 1995). Some of them require complex syntactic manipulation of the sentence, such as *wh*-extraction, and are therefore too difficult to apply automatically. We extend our previous work (Merlo and Leybold, 2001) and choose six diagnostics that can be captured by simple corpus counts: head dependence, optionality, iterativity, ordering, copular paraphrase, and deverbal nominalisation. These diagnostics tap into the deeper semantic properties that distinguish arguments from adjuncts.

Head Dependence Arguments depend on their lexical heads, because they form an integral part of the phrase. Adjuncts do not. Consequently, PP-arguments can only appear with the specific verbal or nominal head by which they are lexically selected, while PP-adjuncts can co-occur with a far greater range of different heads than arguments (Pollard and Sag, 1987, 136), as illustrated in the example sentences below (PP-argument in bold).

- a) a man/woman/scarecrow with gray hair
- b) a student/*punk/*watermelon **of physics**

We capture this insight by measuring the dispersion of the distribution over the different verbs or nouns that co-occur with a given PP in a corpus. We expect adjunct PPs to have higher dispersion than argument PPs. Differently from our previous simpler implementation (Merlo and Leybold, 2001), we use entropy as a measure of the dispersion of the distribution, as indicated in (1) (*h* indicates the noun or verb head to which the PP is attached).

$$H(PP) = -\sum_i p(h_i) \log_2 p(h_i) \quad (1)$$

Optionality In most cases of verb attachment,¹ PP-arguments are obligatory elements of a given sentence whose absence leads to ungrammaticality, while adjuncts do not contribute to the semantics of any particular verb, hence they are optional, as illustrated in the following examples:²

- a) John put the book **in the room**
- b) *John put the book
- c) John saw/read the book in the room
- d) John saw/read the book

The notion of optionality can be captured by the conditional probability of a PP given a particular verbal head, $P(PP|v)$.

¹We do not compute this measure for nominal heads as PP are always optional when governed by a nominal head.

²Notice that this diagnostics can only be interpreted as a statistical tendency, and not as a strict test, because not all arguments are obligatory (but all adjuncts are indeed optional). The best known descriptive exceptions to the criterion of optionality are the class of so-called object-drop verbs (Levin, 1993) and, arguably, instrumental verbs (Schütze, 1995). While keeping these exceptions in mind, we maintain optionality as a valid diagnostic here.

Iterativity and Ordering Arguments cannot be iterated and they must be adjacent to the selecting lexical head. Neither of these two restrictions apply to adjuncts, as illustrated in the examples below.

- a) *Chris rented the gazebo **to girls, to boys**
- b) Kim met Sandy in Baltimore in the hotel lobby in a corner

Thus, the probability of a PP being an adjunct can be approximated as the probability of its occurrence in second position in a sequence of PPs, as indicated in (2).

$$P(ADJ|(PP)_1) \approx P(PP)_2 \quad (2)$$

Copular Paraphrase The diagnostic of copular paraphrase is specific to the distinction of NPs arguments and adjuncts (Schütze, 1995, 103). NP arguments cannot be paraphrased by a copular relative clause (examples b and b'), while adjuncts can (examples a and a').

- a) a man from Paris
- a') a man who was from Paris
- b) the weight of the cow
- b') *the weight that was of the cow

Thus, the probability that a PP is an adjunct can be approximated by the probability of its occurrence following a copular verb, such as *be*, *become*, *appear*, *seem*, *remain* (Quirk et al., 1985), as indicated in (3).

$$P(ADJ|(PP)) \approx P(\text{copular}_v < PP) \quad (3)$$

Deverbal Nouns This diagnostic is based on the observation that PPs following a deverbal noun are likely to be arguments. This diagnostic can be captured by a probability indicators function, that assigns probability 1 of being an argument to PPs following a deverbal noun and 0 otherwise.

In conclusion, the different properties of arguments and adjuncts can be reduced to surface indicators, which can be estimated by appropriate corpus counts.

4 Experiments

The success and generality of corpus-based classifier induction rests in large part on the accurate

estimation of the feature vectors used for training. We explain the details of our methodology below.³

4.1 The Materials

Corpora We construct two corpora comprising examples of PP sequences. A PP is a preposition and head noun sequence. One corpus contains data encoding information for attachment of single PPs in the form of four head words (verb, object noun, preposition and PP-internal noun) for each instance of PP attachments found in the corpus. We also create an auxiliary corpus of sequences of two PPs, where each data item consists of verb, direct object and the two following PPs. This corpus is only used to estimate the feature Iterativity. All the data was newly extracted from the Penn Tree-bank. Our goal was to create a more comprehensive and possibly more accurate corpus than existing ones (Merlo et al., 1997; Collins and Brooks, 1995). To improve coverage, we extracted all cases of PPs following transitive, and intransitive verbs and following nominal phrases. We include also passive sentences and sentences containing a sentential object. To improve accuracy, attention was paid not to extract overlapping data, contrary to counts in previous corpora, where multiple PP sequences were counted more than once, each time as part of a different structural configuration.

The Counts The linguistic diagnostics illustrated in the previous section are approximated by corpus counts based on the extracted tuples.

Head dependence is approximated by the entropy of the distribution of the verb or noun heads for each PP, as indicated in (4).

$$H(PP) \approx -\sum_i \frac{C(h_i)}{\sum_i C(h_i)} \log_2 \frac{C(h_i)}{\sum_i C(h_i)} \quad (4)$$

We implement also some more general variants, where PP-internal nouns and head nouns are replaced by their WordNet 1.7 class (Miller et al., 1990). Polysemous nouns are disambiguated by selecting the most frequent WordNet sense.

Optionality is captured by the conditional probability of a PP given a particular verbal head, as indicated in (5).

$$P(PP|v) \approx \frac{C(v, PP)}{C(v)} \quad (5)$$

Analogously to the measure of head dependence, optionality is also measured in general variants that rely on verb and noun classes, based on WordNet 1.7.

Iterativity and ordering are approximated by collecting counts indicating the proportion of cases in which a given PP in first position had been found in second position in a sequence of multiple PPs over the total of PPs in second position, as indicated in (6).

$$P(ADJ|(PP)_1) \approx \frac{C(PP)_2}{\sum_i C(PP_i)_2} \quad (6)$$

The problem of sparse data here is especially serious, because of the small frequencies of multiple PPs. We estimate this measure in a single variant using a backed-off estimation (Katz, 1987), where we replace lexical items by their WordNet classes.

Copular paraphrase is captured by calculating the proportion of times a given PP follows a copular verb over all the times it appears following any verb.

$$P(ADJ|(PP)) \approx \frac{C(\text{copular}_v < PP)}{\sum_i C(v_i < PP)} \quad (7)$$

This measure is an approximation, since we don't know whether the copular verb is indeed in a relative clause or not. Here again, we back-off to the noun classes of the PP-internal noun, to address the problem of sparse data.

The diagnostic of deverbal nouns is implemented as a binary feature that simply indicates if the PP follows a deverbal noun or not. Deverbal nouns are identified by inspecting their morphology (Quirk et al., 1985). The suffixes that can combine with verb bases to form deverbal nouns are *-ant* (inhabitant), *-ee* (appointee), *-er*, or (singer), *-age* (breakage), *-al* (refusal), *-ion* (exploration), *-sion* (invasion), *-ing* (building), *-ment* (arrangement).

In conclusion, the linguistic diagnostics can be approximated by simple statistical indicators estimated in a sufficiently large corpus. Once the counts are collected they constitute the input to an automatic classifier.

³For more detail on the features and experiments, please see (Esteve-Ferrer and Merlo, 2002).

-CLR	dative object if dative shift not possible(e.g. <i>donate</i>); phrasal verbs; predication adjuncts
-DTV	dative object if dative shift possible (e.g. <i>give</i>)
-BNF	benefactive (dative object of <i>for</i>)
-PRD	non VP predicates
-PUT	locative complement of <i>put</i>
-LGS	logical subjects in passives
-DIR	direction and trajectory
-LOC	location
-MNR	manner
-PRP	purpose and reason
-TMP	temporal phrases

Figure 1: Grammatical function and semantic tags that involve PP constituents in the PTB

The Target Attribute Since we are planning to use a supervised learning method, we need to label each example with a four-valued target attribute (the values are Narg, Nadj, Varg, Vadj). The Penn Treebank annotation does not explicitly make the distinction between arguments and adjuncts. Information about this difference then must be gleaned from the semantic and function tags that have been assigned to the nodes (Bies et al., 1995). Figure 1 illustrates the tags that involve PP constituents. Based on the guidelines (Marcus et al., 1994, 4),(Bies et al., 1995, 12), inspection of the actual annotation in the Tree bank, and discussions in the literature (Quirk et al. 1985, sections 8.27-35, 15.22, 16-48), we mapped PPs into arguments and adjuncts as follows: Adjuncts: All PPs tagged with a semantic tag (DIR, LOC, MNR, PRP, TMP). Arguments: All untagged PPs or PPs tagged with CLR, PUT, DTV, BNF, PRD or LGS.

4.2 The Method

The Input Data Each input vector represents an instance of a PP attachment, which could be both noun or verb attached, either as an argument or as an adjunct.⁴ Each vector contains 20 training features. They comprise the four lexical heads and their WordNet classes, all the different variants of the implementation of the diagnostics, and one 4-valued target feature, indicating the type of attachment.

⁴Notice however that the learning features were calculated on all the instances described above, so in practice we use both the unambiguous cases and ambiguous cases in the estimation of the features of the ambiguous cases.

Experimental Settings We use the C5.0 Decision Tree Induction Algorithm (Quinlan, 1992), applied to a training and testing corpus containing 13906 exemplars, of which we used 90% for training and 10% for testing. The test sets were selected by stratified sampling. (Nine samples were created.)

Clearly, to classify PPs into four classes, we have two options: we can construct a single four-class classifier or we can build a sequence of binary classifiers. The discrimination between noun and verb attachment can be performed first, and then further refined into attachment as argument or adjunct, performing the 4-way classification in two steps. The two-step approach would be the natural way of extending current PP attachment disambiguation methods to the more specific 4-way attachment we propose here. However, based on exploratory data analysis and general wisdom in machine learning, there is reason to believe that it is better to solve the 4-way classification problem directly rather than first solving a more general problem and then specialise the classification.

To test these expectations, we performed both kinds of experiments: a direct 4-way classification experiment, and a two-step classification experiment, to investigate which of the two methods is better. The direct four-way classification uses the attributes described above to build a single classifier. For comparability, we created a two-step experimental setup as follows. We created three binary classifiers. The first one performs the noun-verb attachment classification. Its learning features comprise the four lexical heads and their WordNet classes. We also train two classifiers that learn to distinguish arguments from adjuncts. One classifier is trained only on verb-attachment exemplars and uses only the verb attachment related features. The third classifier is trained only on noun-attachment exemplars, and utilises only the noun attachment related features. The test data is first given to the noun-verb attachment classifier. Then, the test examples classified as verbs are given to the verb argument-adjunct classifier, and the test examples classified as nouns are given to the noun argument-adjunct classifier. Thus, this cascade of classifiers performs the same task as the 4-way classifier, but it does so in two passes.

Task	% Accuracy (% Error reduction)			
	Base	All fts	Bst+w	Bst+w+c
2+2	65.3	70.9(16)	69.9(13)	71.1(17)
4way	66.2	72.7(19)	73.9(23)	73.9(23)

Table 1: Percent accuracy (percent error reduction) using combination of features in the two different experimental settings

Each binary classifier reaches good or even the state of the art accuracy for these tasks. Specifically, the classifier disambiguating the noun-verb attachment reaches an accuracy of 80.2% (baseline 71.6% using only the preposition) using information about argumenthood and 77.2% if the decision tree induction is performed exclusively on the basis of words and classes. The classifier that distinguishes verb arguments from verb adjuncts performs at an accuracy of 81.1% (baseline 72.3%), while the discrimination of noun arguments from noun adjuncts reaches an accuracy of 89.8% (the baseline is already very high, 88.4%).

5 Results and Discussion

Tables 1 reports the classification accuracy of the two comparative experiments performed. All reported numbers are averages of accuracies on 9 different data samples. The first column reports the baseline accuracy for both tasks, calculated by performing the classification using only the feature *preposition*. The second column shows the results of an experiment where all the features described above were used for the classification. The third column reports the results in which only the lexical heads and most effective features – determined experimentally – were used. The fourth column reports the results in which the lexical heads and their classes together with the most effective features were used. For the two-step classification experiment, we determined the best feature for each classifier independently. We observe that the 4-way classification is better in both cases, even if the two-step method decomposes the problem in simpler tasks.

The degradation in performance of the two-step approach compared to the 4-way classification is not unexpected and is likely due to the inappropriate underlying assumptions about the distribu-

tion of the data. A two-step approach assumes that the preferred attachment in the first step subsumes the preferred attachment in the second step. For example, it assumes that a given PP can first be classified as attached to the verb and then refined as a verb argument. Numerically, this assumption is only verified if the largest proportion of cases over the four possible attachments is a subset of the larger proportion of cases in the disambiguation between noun and verb attachment. In general, this assumption holds if the distribution of cases is very skewed. But in some case it does not hold. For example, consider a hypothetical ambiguous PP sequence, distributed across the four possible outcomes with the following proportions: N-arg .25; N-adj .35; V-arg .4; V-adj 0. As can be easily seen, a two-step approach relying directly on the proportion of cases in the data would classify this sequence as an NP attachment at first (.25 + .35 = .6 > .4), assigning it to the N-adj class in a second step. However, the most likely assignment in a 4-way classification, and the correct one, is that the PP is an argument of the verb.

Detailed data analysis of our corpus indicates that some n-way ambiguities exist and that in these cases even distributions of attachments, though rare, do arise in practice. Multiple ambiguities arise, not at the level of word sequences, but at the level of the abstract representations used to cope with sparse data. For example, they arise for PPs represented as a preposition and the semantic class of the PP-internal noun. For such abstract representations, one can find a few instances of the unskewed distribution illustrated above. For these cases, a two-step approach would favour the wrong solution. Consider the case of the PP consisting of the preposition *at* and the WordNet class 14 (group), whose relative frequencies in the Penn Treebank are N-arg .12, N-adj .41, V-arg 0, V-adj .47. A two-step approach would first choose a noun attachment, and further refine the choice to adjunct, while the most frequent case is the verb adjunct attachment case. Another similar example, is the case of the PP consisting of the preposition *from* and the WordNet class 4 (artifact). In this instance, a two-step approach would favour a verb attachment at first, further refined into argument attachment, while the preferred attachment is

Task	F-scores (%)			
	Nadj	Narg	Vadj	Varg
4-way	47	86	63	56
2+2	51	86	54	47

Table 2: Percent F-scores using best features in the two different experimental settings

as noun argument. (Relative frequencies are N-arg .47, N-adj 0, V-arg .44, V-adj .09.)

A more refined analysis of the errors of our classifiers confirms this interpretation. Table 2 shows the F-scores for a sample used in the experiment reported in the fourth column of Table 1. As can be seen, the 4-way classification is a little worse for noun attachments (due to worse recall of noun adjuncts) but better for the verb attachments. The distribution of the errors reveals that improvements can be observed in distinguishing noun arguments and verb arguments – clearly an instance in which the actual attachment site does not subsume the assigned attachment. This result confirms that it is misleading to formalise PP attachment as a binary problem, assuming that a distinction between arguments and adjuncts can be performed later, if necessary, without loss of accuracy.

The 4-way classification also reveals a little improvements in the discrimination of verb arguments from verb adjuncts (especially fewer verb arguments misclassified as adjuncts). Improvements in the precision of verb arguments is particularly relevant, as this class of attachments is crucial in supporting further language processing tasks, which usually require precise knowledge of a verb’s subcategorization frame.

In conclusion, these results show that good accuracy on a fine-grained and informative classification of PPs can be achieved using corpus counts. Moreover, classification of PPs performed with a single classifier is more accurate than sequencing binary classifiers: by tackling the 4-way discrimination problem directly, the approach does not rely on any assumption on the distribution of the data, and thereby reaches better accuracy. Finally, a result of particular interest is the improved discrimination of verb arguments from verb adjuncts compared to a two-step approach, a promising result

for all the numerous NLP tasks and applications that rely on the correct identification of a verb’s subcategorization frame.

6 Related Work

As far as we are aware, this is the first attempt to integrate the notion of argumenthood in a more comprehensive formulation of the problem of disambiguating the attachment of PPs. Hindle and Rooth (1993) mention the interaction between the structural and the semantic factors in the disambiguation of a PP, indicating that verb adjuncts are the most difficult. We confirm their finding that noun arguments are more easily identified, while verb complements (either arguments or adjuncts) are more difficult.

Few previous pieces of work attempt to distinguish arguments from adjuncts automatically (Buchholz, 1999; Merlo and Leybold, 2001). We extend here (Merlo and Leybold, 2001) by elaborating more learning features, refining all the counting methods and extending the method to noun attachment, which had not been considered before, thus validating and extending the approach. The current work on binary argument-adjunct classifiers compares favourably to the only other comparable study on this topic (Buchholz, 1999). Buchholz reports an accuracy of 77% for the argument-adjunct distinction of PPs, to be compared to our 81% and 89% for verb and noun attachments respectively. Buchholz considers all types of attachment sites, not just verbs and nouns. Recently (Villavicencio, 2002) has explored the performance of an argument identifier, developed in the framework of a model of child language learning. The approach is not directly comparable, as it is not entirely corpus-based (the input to the algorithm is an impoverished logical form), and the evaluation is on a smaller scale than the present work. Other pieces of work address the current problem in the larger perspective of distinguishing arguments from adjuncts for subcategorization acquisition (Korhonen, 2002; Aldeza et al., 2002). Our work confirms the results reported in (Korhonen, 2002), which indicate that using word classes improves the extraction of subcategorisation frames.

7 Conclusions

We have proposed a reformulation of the problem of PP attachment as a 4-way disambiguation problem, arguing that what is needed in interpreting prepositional phrases is knowledge about both the structural attachment site – the traditional noun-verb attachment distinction – and the nature of the attachment – the distinction of arguments from adjuncts. Practically, we have shown that a 4-way classifier which solves the complete problem directly performs better than solving a sequence of binary decisions. Future work lies in further investigating the difference between arguments and adjuncts to achieve even finer-grained classifications and to model more precisely the semantic core of a sentence.

Acknowledgments

This research was made possible by Swiss NSF grant no. 11-65328.01. I would like to thank Eva Esteve Ferrer for her collaboration.

References

- Izaskun Aldezabal, Maxux Aranzabe, Koldo Gojenola, Kepa Sarasola, and Aitziber Atutxa. 2002. Learning argument/adjunct distinction for Basque. In *Procs of the SIGLEX Workshop on Unsupervised Lexical Acquisition*, pages 42–50, Philadelphia, PA, July.
- Ann Bies, M. Ferguson, K.Katz, and Robert MacIntyre. 1995. Bracketing guidelines for Treebank II style. Technical report, University of Pennsylvania.
- Sabine Buchholz. 1999. Distinguishing complements from adjuncts using memory-based learning. ILK, Computational Linguistics, Tilburg University.
- Michael Collins and James Brooks. 1995. Prepositional Phrase Attachment through a Backed-Off Model. In *Procs of the Third Workshop on Very Large Corpora*, pages 27–38, Cambridge, MA.
- Bonnie Dorr. 1997. Large-scale dictionary construction for foreign language tutoring and interlingual machine translation. *Machine Translation*, 12(4):1–55.
- Eva Esteve-Ferrer and Paola Merlo. 2002. Automatic distinction of PP arguments and adjuncts. Technical report, MALA Project 1, University of Geneva.
- Jane Grimshaw. 1990. *Argument Structure*. MIT Press.
- Donald Hindle and Mats Rooth. 1993. Structural ambiguity and lexical relations. *Computational Linguistics*, 19(1):103–120.
- Ray Jackendoff. 1977. *X' Syntax: A Study of Phrase Structure*. MIT Press, Cambridge, MA.
- S.M. Katz. 1987. Estimation of probabilities from sparse data for the language model component of a speech recognizer. *IEEE Transactions on Acoustic, Speech and Signal Processing*, 35(3):400–401.
- Anna Korhonen. 2002. Semantically motivated subcategorization acquisition. In *Procs of the SIGLEX Workshop on Unsupervised Lexical Acquisition*, pages 51–58, Philadelphia, PA, July.
- Beth Levin. 1993. *English Verb Classes and Alternations*. University of Chicago Press, Chicago, IL.
- A. Marantz. 1984. *On the Nature of Grammatical Relations*. MIT Press, Cambridge, MA.
- M. Marcus, G. Kim, A. Marcinkiewicz, R. Macintyre, A. Bies, M. Ferguson, K.Katz, and B. Schasberger. 1994. The Penn Treebank: Annotating argument structure. Technical report, University of Pennsylvania.
- Paola Merlo and Matthias Leybold. 2001. Automatic distinction of arguments and modifiers: the case of prepositional phrases. In *Procs of the Fifth Computational Natural Language Learning Workshop (CoNLL-2001)*, pages 121–128, Toulouse, France.
- Paola Merlo, Matt Crocker, and Cathy Berthouzoz. 1997. Attaching multiple prepositional phrases: Generalized backed-off estimation. In *Procs of the Second Conference on Empirical Methods in Natural Language Processing*, pages 145–154, Providence, RI.
- George Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. Miller. 1990. Five papers on Wordnet. Technical report, Cognitive Science Lab, Princeton University.
- Carl Pollard and Ivan Sag. 1987. *An Information-based Syntax and Semantics*, volume 13. CSLI lecture Notes, Stanford University.
- J. Ross Quinlan. 1992. *C4.5 : Programs for Machine Learning*. Series in Machine Learning, Morgan Kaufmann, San Mateo, CA.
- Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Longman, London.
- Adwait Ratnaparkhi, Jeffrey Reynar, and Salim Roukos. 1994. A Maximum Entropy Model for Prepositional Phrase Attachment. In *Procs of the ARPA Workshop on Human Language Technology*, pages 250–255.
- Carson T. Schütze. 1995. PP Attachment and Argumenthood. *MIT Working Papers in Linguistics*, 26:95–151.
- Bangalore Srinivas and Aravind K. Joshi. 1999. Supertagging: An approach to almost parsing. *Computational Linguistics*, 25(2):237–265.
- Aline Villavicencio. 2002. Learning to distinguish PP arguments from adjuncts. In *Procs of the 6th Conference on Natural Language Learning (CoNLL-2002)*, pages 84–90, Taipei, Taiwan.