Determining the Position of Adverbial Phrases in English

Huayan Zhong and **Amanda Stent**

Computer Science Department

Stony Brook University

Stony Brook, NY 11794, USA

`zhong@cs.sunysb.edu, amanda.stent@stonybrook.edu`

Abstract

In this paper we compare three approaches to adverbial positioning using lexical, syntactic, semantic and sentence-level features. We find that: (a), one- and two-stage classification-based approaches can achieve almost 86% accuracy in determining the absolute position of adverbials; (b) a classifier trained with only syntactic features gives performance close to that of a classifier trained with all features; and (c) a surface realizer incorporating a two-stage classifier for adverbial positioning as the second stage gives improvements of at least 10% in simple string accuracy over a baseline realizer for sentences containing adverbials.

1 Introduction

The job of a *surface realizer* is to transform an input semantic/syntactic form into a sequence of words. This task includes word choice, and word and constituent ordering. In English, the positions of required elements of a sentence, verb phrase or noun phrase are relatively fixed. However, many sentences also include *adverbials* whose position is not fixed (Figure 1). There may be several appropriate positions for an adverbial in a particular context, but other positions give output that is non-idiomatic or disfluent, ambiguous, or incoherent.

Some computational research has included models for adjunct ordering (e.g. (Ringger et al., 2004; Marciniak and Strube, 2004; Elhadad et al., 2001)). However, this is the first computational study to look specifically at adverbials. Adverbial positioning has long been studied in linguistics (e.g. (Keyser, 1968; Allen and Cruttenden, 1974; Ernst, 1984; Haider, 2000)). Most linguistic research focuses on whether adverbial placement is functional or semantic in nature. However, Costa (2004) takes a more flexible feature-based approach that uses: lexical features (e.g. phonological shape, ambiguity of meaning, categorical status); syntactic features (e.g. possible adjunction sites, directionality of adjunction, domain of modification); and information structure features (e.g. focus, contrast). We decided to evaluate Costa's approach computationally, using features automatically extracted from an annotated corpus.

In this paper, we compare three approaches to adverbial positioning: a simple baseline approach using lexical and syntactic features, and one- and two-stage classification-based approaches using lexical, syntactic, semantic and sentence-level features. We apply these approaches in a hybrid surface realizer that uses a probabilistic grammar to produce realization alternatives, and a second-stage classifier to select among alternatives. We find that: (a) One- and two-stage classification-based approaches can achieve almost 86% accuracy in determining the absolute position of adverbials; (b) A classifier trained with only syntactic features gives performance close to that of a classifier trained with all features; and (c) A surface realizer using a two-stage classifier for adverbial positioning can get improvements of at least 10% in simple string accuracy over a baseline realizer for sentences containing adverbials.

As well as being useful for surface realization, a model of adverbial ordering can be used in machine translation (e.g. (Ogura et al., 1997)), language learning software (e.g. (Leacock, 2007; Burstein et al., 2004)), and automatic summarization (e.g. (Elhadad et al., 2001; Clarke and Lapata, 2007; Madnani et al., 2007)).
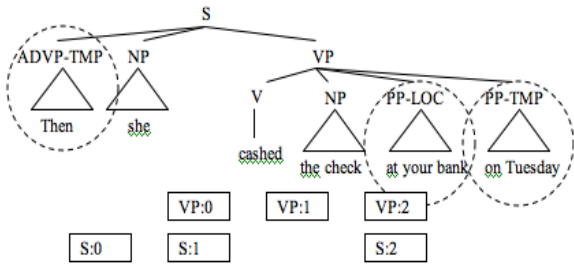
Figure 1: Example syntax tree for *Then she cashed the check at your bank on Tuesday* with adverbials circled and possible VP and S adverbial positions in squares.

2 Data and Features

From the sentences in the Wall Street Journal (WSJ) and Switchboard (SWBD) sections of the Penn Treebank III (Marcus et al., 1999), we extracted all NP, PP and ADVP phrases labeled with the adverbial tags -BNF, -DIR, -EXT, -LOC, -MNR, -PRP, -TMP or -ADV. These phrases mostly modify S constituents (including RRC, S, SBAR, SBARQ, SINV, SQ), VP constituents, or NP constituents (including NP and WHNP), but also modify other adjuncts (PP, ADJP or ADVP) and other phrase types (FRAG, INTJ, LST, NAC, PRT, QP, TOP, UCP, X).

Corpus	Number of adverbials of type:		
	PP-ADVP	NP-ADVP	ADVP
WSJ	36128	10587	13700
SWBD	12231	5405	17193

Table 1: Adverbials in the Penn Treebank III

For each adverbial, we automatically extracted lexical, syntactic, semantic and discourse features. We included features similar to those in (Costa, 2004) and from our own previous research on prepositional phrase ordering (Zhong and Stent, 2008). Due to the size of our data set, we could only use features that can be extracted automatically, so some features were approximated. We dropped adverbials for which we could not get features, such as empty adverbials. Tables 1 and 2 summarize the resulting data. A list of the features we used in our classification experiment appears in Table 3. We withheld 10% of this data for our realization experiment.

3 Classification Experiment

Our goal is to determine the position of an adverbial with respect to its siblings in the phrase of which it

Adverbial	Data Set	
Type	WSJ	SWBD
S	8196	5144
VP	29734	22845
NP	12985	2071
PP/ADJP/ADVP	1739	987
Other	297	686

Table 2: Adverbials in the Penn Treebank III

is a part. An adverbial may have non-adverbial siblings, whose position is typically fixed. It may also have other adverbial siblings. In the sentence in Figure 1, *at your bank* has one adverbial and two non-adverbial siblings. If this adverbial were placed at positions VP:0 or VP:1 the resulting sentence would be disfluent but meaningful; placed at position VP:2 the resulting sentence is fluent, meaningful and idiomatic. (In this sentence, both orderings of the two adverbials at position VP:2 are valid.)

3.1 Approaches

We experimented with three approaches to adverbial positioning.

Baseline Our baseline approach has two stages. In the first stage the position of each adverbial with respect to its non-adverbial siblings is determined: each adverbial is assigned the most likely position given its lexical head and category (PP, NN, ADVP). In the second stage, the relative ordering of adjacent adverbials is determined in a pairwise fashion (cf. (Marciniak and Strube, 2004)): the ordering of a pair of adverbials is assigned to be the most frequent in the training data, given the lexical head, adverbial phrase type, and category of each adverbial.

One-stage For our one-stage classification-based approach, we determine the position of all adverbials in a phrase at one step. There is one feature vector for each phrase containing at least one adverbial. It contains features for all non-adverbial siblings in realization order, and then for each adverbial sibling in alphabetical order by lexical head. The label is the order of the siblings. For example, for the S-modifying adverbial in Figure 1, the label would be 2_0_1, where 0 = "she", 1 = "cashed" and 2 = "Then". If there are n siblings, then there are $n!$ possible labels for each feature vector, so the performance of this classifier by chance would be .167 if each adverbial has on average three siblings.

Type	Features
lexical	preposition in this adverbial and in adverbial siblings 0-4; stems of lexical heads of this adverbial, its parent, non-adverbial siblings 0-4, and adverbial siblings 0-4; number of phonemes in lexical head of this adverbial and in lexical heads of adverbial siblings 0-4; number of words in this adverbial and in adverbial siblings 0-4
syntactic	syntactic categories of this adverbial, its parent, non-adverbial siblings 0-4, and adverbial siblings 0-4; adverbial type of this adverbial and of adverbial siblings 0-4 (one of DIR, EXT, LOC, MNR, PRP, TMP, ADV); numbers of siblings, non-adverbial siblings, and adverbial siblings
semantic	hypernyms of heads of this adverbial, its parent, non-adverbial siblings 0-4, and adverbial siblings 0-4; number of meanings for heads of this adverbial and adverbial siblings 0-4 (using WordNet)
sentence	sequence of children of S node (e.g. NP VP, VP); form of sentence (declarative, imperative, interrogative, clause-other); presence of the following in the sentence: coordinating conjunction(s), subordinating conjunction(s), correlative conjunction(s), discourse cue(s) (e.g. 'however', 'therefore'), pronoun(s), definite article(s)

Table 3: Features used for determining adverbial positions. We did not find phrases with more than 5 adverbial siblings or more more than 5 non-adverbial siblings. If a phrase did not have 5 adverbial or non-adverbial siblings, NA values were used in the features for those siblings.

Two-stage For our two-stage classification-based approach, we first determine the position of each adverbial in a phrase in relation to its non-adverbial siblings, and then the relative positions of adjacent adverbials. For the first stage we use a classifier. There is one feature vector for each adverbial. It contains features for all non-adverbial siblings in realization order, then for each adverbial sibling in alphabetical order by lexical head, and finally for the target adverbial itself. The label is the position of the target adverbial with respect to the non-adverbial siblings. For our example sentence in Figure 1, the label for "Then" would be 0; for "at the bank" would be 2, and for "on Tuesday" would be 2. If there are n non-adverbial siblings, then there are $n + 1$ possible labels for each feature vector, so the performance of this classifier by chance would be .25 if each adverbial has on average three non-adverbial siblings.

For the second stage we use the same second stage as the baseline approach.

3.2 Method

We use 10-fold cross-validation to compute performance of each approach. For the classifiers, we used the J4 decision tree classifier provided by Weka[1]. We compute correctness for each approach as the percentage of adverbials for which the approach outputs the same position as that found in the original

human-produced phrase. (In some cases, multiple positions for the adverbial would be equally acceptable, but we cannot evaluate this automatically.)

3.3 Results

Our classification results are shown in Table 4. The one- and two-stage approaches both significantly outperform baseline. Also, the two-stage approach outperforms the one-stage approach for WSJ.

The decision trees using all features are quite large. We tried dropping feature sets to see if we could get smaller trees without large drops in performance. We found that for all data sets, the models containing only syntactic features perform only about 1% worse for one-stage classification and only about 3% worse for two-stage classification, while in most cases giving much smaller trees (1015 [WSJ] and 972 [SWBD] nodes for the one-stage approach; 1008 [WSJ] and 877 [SWBD] for the two-stage approach). This is somewhat surprising given Costa's arguments about the need for lexical and discourse features; it may be due to errors introduced by approximating discourse features automatically, as well as to data sparsity in the lexical features.

There are only small performance differences between the classifiers for speech and those for text.

4 Realization Experiment

To investigate how a model of adverbial positioning may improve an NLP application, we incorpo-

---

[1]We experimented with logistic regression and SVM classifiers; the decision tree classifier gave the highest performance.

Approach	Tree size	Classification accuracy	SSA
WSJ			
baseline	n/a	45.98	75.1
one-stage	6519	84.43	82.2
two-stage	1053	86.27	85.1
SWBD			
baseline	n/a	41.48	61.3
one-stage	4486	85.13	74.5
two-stage	3707	85.01	73.1

Table 4: Performance of adverbial position determination

rated our best-performing models into a surface realizer. We automatically extracted a probabilistic lexicalized tree-adjoining grammar from the whole WSJ and SWBD corpora minus our held-out data, using the method described in (Zhong and Stent, 2005). We automatically re-realized all adverbial-containing sentences in our held-out data (10%), after first automatically constructing input using the method described in (Zhong and Stent, 2005).

We compute realization performance using simple string accuracy (SSA)[2]. Realization performance is reported in Table 4. Both classification-based approaches outperform baseline, with the two-stage approach performing best for WSJ with either metric (for SWBD, the classification-based approaches perform similarly).

5 Conclusions and Future Work

In this paper, we tested classification-based approaches to adverbial positioning. We showed that we can achieve good results using syntactic features alone, with small improvements from adding lexical, semantic and sentence-level features. We also showed that use of a model for adverbial positioning leads to improved surface realization. In future work, we plan a human evaluation of our results to see if more features could lead to performance gains.

6 Acknowledgments

---

[2]Although in general we do not find SSA to be a reliable metric for evaluating surface realizers, in this case it is valid because lexical selection is done already; only the positions of adverbials will generally be different.

References

D. Allen and A. Cruttenden. 1974. English sentence adverbials: Their syntax and their intonation in British English. *Lingua*, 34:1–30.

J. Burstein, M. Chodorow, and C. Leacock. 2004. Automated essay evaluation: The Criterion online writing service. *AI Magazine*, 25(3):27–36.

J. Clarke and M. Lapata. 2007. Modelling compression with discourse constraints. In *Proceedings of EMNLP/CoNLL*.

J. Costa. 2004. A multifactorial approach to adverb placement: assumptions, facts, and problems. *Lingua*, 114:711–753.

M. Elhadad, Y. Netzer, R. Barzilay, and K. McKeown. 2001. Ordering circumstantials for multi-document summarization. In *Proceedings of BISFAI*.

Thomas Ernst. 1984. *Towards an integrated theory of adverb position in English*. Ph.D. thesis, Indiana University, Bloomington, Indiana.

H. Haider. 2000. Adverb placement. *Theoretical linguistics*, 26:95–134.

J. Keyser. 1968. Adverbial positions in English. *Language*, 44:357–374.

C. Leacock. 2007. Writing English as a second language: A proofreading tool. In *Proceedings of the Workshop on optimizing the role of language in technology-enhanced learning*.

N. Madnani, D. Zajic, B. Dorr, N. F. Ayan, and J. Lin. 2007. Multiple alternative sentence compressions for automatic text summarization. In *Proceedings of the Document Understanding Conference*.

T. Marciniak and M. Strube. 2004. Classification-based generation using TAG. In *Lecture Notes in Computer Science*, volume 3123/2004. Springer Berlin/Heidelberg.

M. Marcus, B. Santorini, M. Marcinkiewicz, and A. Taylor. 1999. Treebank-3. Available from the Linguistic Data Consortium, Catalog Number LDC99T42.

K. Ogura, S. Shirai, and F. Bond. 1997. English adverb processing in Japanese-to-English machine translation. In *Seventh International Conference on Theoretical and Methodological Issues in Machine Translation*.

E. Ringger, M. Gamon, R. Moore, D. Rojas, M. Smets, and S. Corston-Oliver. 2004. Linguistically informed statistical models of constituent structure for ordering in sentence realization. In *Proceedings of COLING*.

H. Zhong and A. Stent. 2005. Building surface realizers automatically from corpora using general-purpose tools. *Proceedings of UCNLG*.

H. Zhong and A. Stent. 2008. A corpus-based comparison of models for predicting ordering of prepositional phrases. *In submission*.