AUTOMATIC LABELING OF CONTRASTIVE WORD PAIRS FROM SPONTANEOUS SPOKEN ENGLISH

Leonardo Badino, Robert A.J. Clark

Centre for Speech Technology Research
University of Edinburgh, Edinburgh, UK

ABSTRACT

This paper addresses the problem of automatically labeling contrast in spontaneous spoken speech, where contrast here is meant as a relation that ties two words that explicitly contrast with each other. Detection of contrast is certainly relevant in the analysis of discourse and information structure and also, because of the prosodic correlates of contrast, could play an important role in speech applications, such as text-to-speech synthesis, that need an accurate and discourse context related modeling of prosody. With this prospect we investigate the feasibility of automatic contrast labeling by training and evaluating on the Switchboard corpus a novel contrast tagger, based on Support Vector Machines (SVM), that combines lexical features, syntactic dependencies and WordNet semantic relations.

Index Terms— spoken language understanding, information structure, contrast, syntactic dependencies, WordNet, support vector machines

1. INTRODUCTION

The concept of contrast plays an important role in many spoken language technologies, ranging from spoken language understanding to speech synthesis. According to the observation point one looks at it, contrast can be seen as: a) a discourse relation that ties discourse elements; b) a concept of information structure that makes a word (or a phrase) salient by comparing it with other word(s) available from the discourse context; c) a linguistic concept often prosodically marked.

Given the broad meaning of contrast, the different discourse scenarios invoking it, the poor availability of corpora annotated with categories of contrast, and our main research interest of investigating the role of contrast in prosodic prominence modeling for text-to-speech applications, we decided to focus on one aspect/category of contrast only: an information structure relation that links two semantically related words that explicitly contrast with each other. This category, along with the contrast categories: *correction*, *subset*, *adverbial* and *answer*, was used to manually annotate information structure in a section of the Switchboard corpus ([1]). (1) is an example (of what hereafter we will simply call *contrast* and is called *contrastive kontrast* in [1]):

(1) **We** seemed to be unfairly doing all the **cooking** and **they** were doing all the **enjoying**.

where "they" contrasts "we" and "enjoying" contrasts "cooking".

In the literature there are only few works on the automatic detection of contrastive information and they differ from ours for the type of contrast they detect and/or the corpus they use. Most of them use acoustic features among their training features whereas we are interested on the textual patterns of contrastiveness only.

In [2] Zhang et al. automatically label *symmetric contrast* by training their labeler on a limited-domain intelligent tutoring system corpus and using a combination of acoustic features (F0, duration, energy and spectral balance cepstral coefficients), Part-Of-Speech, and a semantic similarity measure computed by using both WordNet semantic lexicon and corpora statistics. *Symmetric contrast* differs from our *contrast* but the two concepts overlap.
In [3] a subsection of the Switchboard corpus annotated by [1] is used to detect all the annotated contrast categories. The detector looks at the acoustic properties, POS and probability of being accented, of each single word, in order to label it as contrastive or not, thus no distinction is made among the different contrast categories, and no relation is detected (which word contrasts with which word?).

In this paper we show through examples and experimental evidence that several syntactic and semantic patterns of *contrast* can be consistently recognized by enriching the set of features proposed in previous works with lexical, and deeper syntactic and semantic features.

2. DATA PREPARATION

The Switchboard corpus ([4]) consists of 2430 spontaneous phone conversations (average six minutes) in American English. A third of it is syntactically annotated (POS and grammatical constituents) as part of the Penn Treebank ([5]) and a

Thanks to EPSRC and ICCS for funding.

subsection (146 dialogues) of that third is annotated with information structure ([1]). In the following sections (2.1 and 2.2) we describe how we built the training data for our *contrast* tagger and what kind of restrictions we imposed on the examples of *contrast*. All the syntactic information we used to train our tagger come from the PennTreebank manual annotation. We made this choice in order to explore the potential of our tagger as much as possible independently from the errors of the modules it receives information from.

2.1. Data collection

Before merging the syntactic and the information structure annotations we converted the constituent format in the Penn Treebank into dependency trees using the Penn2Malt converter ([6]). Since the PennTreebank constituent annotation for Switchboard uses slightly different (and not yet standardly held) conventions from whose presupposed by the Penn2Malt converter we had to support the converter with some additional scripts. However, because of problems we encountered in the conversion process we had to remove 54 (out of 146) dialogues. For each remaining dialogue all the word senses (according to the WordNet senses set) were disambiguated using the WordNet::SenseRelate Perl module ([7]).

2.2. Data pruning

Not all the sentences of the 92 dialogues and not all the examples of *contrast* were used to train and evaluate our tagger. First, for reasons of computational efficiency, which will be clear in the next sections, we decided to only consider *contrast* relations that occurred within a single dependency tree (i.e. a single sentence, whose boundaries were given by the PennTreebank constituent annotation).Then, we removed all the sentences that did not contain *contrast* (within a single dependency tree). Note that the Subsequently, we decided to consider *contrast* relations linking single words only (as in example (1)) so sentences only containing *contrast* linking phrases of more than one word were removed. This decision was dictated by our aim of focusing on the patterns of contrastiveness without paying attention, for the moment, to the scope of contrastive elements which is a hard and still on debate issue. We also decided, in order to make the tagger's task a bit simpler, to only look at *contrasts* that linked words having the same broad POS: noun, verb, adjective, adverb, pronoun, cardinal number, other. This pruning regarded a very small number of *contrasts*. Since our *contrast* tagger relies on textual features only and does not look at the discourse context outside the sentence containing *contrast*, we removed: 1) all *contrast* relations that we could not identify by simply looking at text, and that had been labeled only because they were prosodically signaled; 2) all *contrasts* activated by discourse items outside the sentence. In this last pruning step some decisions were hard to make, since *contrast* resulted in a combination of prosodic, syntactic, semantic and pragmatic clues. When we were not sure about keeping or removing the sentences containing the problematic *contrast* we kept it.

Note that cases where *contrast* was neither syntactically or semantically determined but only pragmatically determined (as in example (2)) were not removed.

(2) as a **westerner** in **India** ... I was often surprised ...

The final data we used to generate positive and negative examples of contrastive word pairs consisted of 254 sentences containing at least one *contrast* that was not pruned out.

W1	W2	Example value
we	they	+1
seemed	be	-1
seemed	doing	-1
	...	
cooking	enjoying	+1
	...	

Fig. 1. Examples values generation. *Examples values generation from the sentence: WE/PRP seemed/VBD to/TO be/VB unfairly/RB doing/VBG all/PDT the/DT COOKING/NN and/CC THEY/PRP were/VBD doing/VBG all/PDT the/DT ENJOYING/NN. The example value is positive (+1) if W1 and W2 are linked by a* contrast *and negative (-1) otherwise.*

2.3. Examples extraction

For each sentence both positive and negative examples were extracted as shown in Fig.1: all word pairs having the same broad POS were extracted and then assigned a +1 if the two words were linked by *contrast* or a -1 otherwise. An example consisted of its positive or negative value and a sequence of training features. The fact that the computation of some features requires a considerable computational effort (but still reasonable for real time applications) and sentences can be 80 words long or more explains our decision of limiting the *contrast* relations to those occurring within a single sentence.

3. FEATURES EXTRACTION

The features we extracted can be grouped into three categories: lexical features, syntactic features and semantic features. For sake of simplicity hereafter we will refer to the two words of each word pair as W1 and W2, where W1 precedes W2 in the sentence. We also introduce the concept of subsentence: a sub-sentences is a part of a sentence that refers to "verb-dominated" sub-trees. "Verb-dominated" sub-trees are parts of the dependency tree that have a verb as a root. For example, in the sentence:

(3) So well... **you** take this subject much more personally than **I** do, I suppose.

"So well... you take this subject much more personally than", "I do" and "I suppose" are all sub-sentences dominated by the verbs "take", "do" and "suppose" respectively.

3.1. Lexical features

Examples of lexical features are: all CAP words between W1 and W2, first CAP word preceding W1, first CAP word preceding W2, first two CAP words preceding W1, first two CAP words preceding W2. CAP words are conjunctions, adverbs and prepositions.
These features were intended to capture single words or bigrams that activate *contrast*, like, for example, the bigram "rather than" in the sentence:

(4) So she's going to **sell** it rather than **trade** it in.

A feature to measure the degree of textual parallelism between the two subsentences containing W1 and W2 (when W1 and W2 belonged to two different subsentences) was also used since textual parallelism can be a clue of *contrast* as in example (1) and in the following example:

(5) ... let's do **this** way, let's do **that** way ...

The parallelism (normalized) score was computed using the Wagner & Fischer edit distance to compare strings of text as proposed by [8].

3.2. Syntactic features

All syntactic features are POS, dependency relations (subject of, object of, etc...) and features derived from both of them. Examples of features derived from POS are the features indicating if W1 is the only word in the sentence having the same broad POS of W2, and the feature indicating if W1 is the closest (in term of words between them) word preceding W2 and having the same broad POS.

The use of deeper than POS syntactic information such as syntactic dependencies (and information related to them) is motivated by the need of identifying syntactic patterns of contrastiveness that can not be identified using POS and lexical features alone. For example knowing that W1 and W2 have the same type of dependency with their heads as in example (3) (both "you" and the first "I" have a "subject of" dependency with "take" and "do" respectively) or that their heads refer to the same item as in example (6), seems to be a necessary (but often not sufficient) information to identify *contrast*.

(6) and, you know, even the **public** schools are behind the **parochial** schools.

In order to improve the detection of parallelism for two words belonging to two different sub-sentences, we also used features indicating if the two sentences had the subject referring to the same item. The same type of features was used for syntactic objects, dominant verbs and predicates.

3.3. Semantic features

Semantic features seem to be a necessary information for our tagger as well, since contrastive words are usually semantically similar in one way but different in another, as in:

(7) and you see **women** going off to work as well as **men**.

The semantic features set consists of features indicating if W1 and W2 were linked by one of the following WordNet semantic relation: hypernyms, antonyms, entails, member-of, part-of, sisters (that is, two words having the same hypernym). We also used the Lin's semantic similarity measures ([9]) applied to WordNet. Semantic relations and similarity were computed using the WordNet::QueryData and WordNet::Similarity ([10]) Perl modules. Since WordNet relations and similarity measures relate to word senses, they were computed in two different way: (1) on the senses (one per word) provided by the word sense disambiguator (see section 2.1); 2) on the first 3 senses (or less if the word had less than 3 senses) of each word, so a maximum of 9 sense pairs were compared. In the latter case, the chosen similarity score was that referring to the sense pairs producing the highest score.

4. EVALUATION

Our *contrast* tagger is a SVM based predictor. We used the SVM-light implementation ([11]) which allowed us to use different types of kernels: linear, polynomial, radial basis function, sigmoid tanh. The training and testing set consisted of 8602 examples: 275 positives and 8327 negatives. The tagger was evaluated using a leave-one-out estimation of accuracy. The polynomial kernel turned out to be the most effective one. Table 1 shows the values of accuracy, and precision and recall for different orders of the polynomial kernel. The quadratic polynomial gave the best results. A possible explanation for the supremacy of the quadratic polynomial is that the non-linear transformation of the data allow to capture dependencies among training features that are often correlate, whereas the linear polynomial is not able to capture such dependencies. However polynomials with higher order seem to overfit the data. The very unbalanced numbers of positive and negative examples induced us to try different values of the SVM-light training parameter *j* that is the ratio between the cost on false negatives and the cost on false positives. $j = 2$ gave the best results. Trying values of *j* higher than 1 was also motivated by the fact that in a few sentences containing *contrast* between two words also *contrast* between phrases occurred, but the training examples extracted from them were labeled as negatives. For example: Debbie - who → -1, More - who → -1, in sentence (8):

(8) ... **I** 'm *Debbie More*, you know, may I ask you *who* **you** are and ..

Features	d	j	Accuracy	Precision	Recall
Baseline			96.80%	0%	0%
All	1	2	97.02%	70.21%	12.00%
All	2	1	96.88%	65.22%	5.54%
All	2	2	**97.19%**	**76.19%**	**17.45%**
All	2	3	97.17%	65.59%	22.18%
All	3	2	97.00%	68.09%	11.64%
No WordNet	2	2	96.95%	61.62%	13.09%
POS+WordNet	2	2	96.98%	58.43%	18.91%

Table 1. Evaluation of the *contrast* tagger. *d is the order of the polynomial kernel. j is the ratio between the cost on false negatives and the cost on false positives. Precision and recall are relative to positive examples. The baseline is a tagger that always labels examples as non-contrastive (-1)*

Limiting the sentences to those containing *contrast* between two words only would have been preferable but such a constraint would have drastically reduced the number of positive examples. Other "false negatives" in the training data were due to: 1) no prosodically prominent *contrast* was not manually annotated; 2) the *contrast* category overlaps with the other categories defined by [1] (subset, answer,etc...).

We believe that these "false negatives" in the training data have affected the performance of the tagger and that consequently its accuracy rates should be considered as the bottom threshold estimation of its actual accuracy.

Another reason of poor recall may reside in the fact that in a few cases the manual annotation limited the scope of the *contrast* relation to a single word pair only even though the *contrast* relation had a larger scope (that is, it actually referred to phrases or even topics).

Concerning the importance of the different feature categories, results reported in the last two rows in table 1 clearly show the benefits arising from the combined used of lexical, semantic and syntactic dependencies related features.

Analyzing the tagger at a sentence level we observed that in most of the cases it was able to detect *contrast* when *contrast* was activated by textual and syntactic parallelism, or/and by phrases that signal comparison such as "rather than", "instead of". We also found out that the identification of semantic relations was very poor and that affected the tagger's accuracy.

5. CONCLUSION

In this paper we propose a novel approach to automatically label contrastive word pairs from spontaneous spoken English. Although the training data contains a small number of positive examples, our tagger is able to consistently identify *contrast* when *contrast* is activated by a strong textual and syntactic parallelism, or by prepositional and adverbial phrases involving comparisons between two items. Nevertheless our tagger achieves a low recall rate that is due both to the difficulty of the task, which requires richer semantic and pragmatic information, and the nature of the training data, which contains several "false negatives".

6. REFERENCES

[1] M. Calhoun, S.and Nissim, M. Steedman, and J. Brenier, "A framework for annotating information structure in discourse," in *Frontiers in Corpus Annotation II: Pie in the Sky, ACL2005 Conference Workshop*, Arbor, Michigan, 2005.

[2] T. Zhang, M. Hasegawa-Johnson, and S.E. Levinson, "Exraction of pragmatic and semantic salience from spontaneous spoken english," *Speech Communication*, vol. 48, 2006.

[3] A. Nenkova and D. Jurafsky, "Automatic detection of contrastive elements in spontaneous speech," in *IEEE workshop on Automatic Speech Recognition and Understanding (ASRU)*, Kyoto, Japan, 2007.

[4] J. Godfrey, E. Holliman, and J. McDaniel, "Switchboard: Telephone speech corpus for research and development," in *Proc. of ICASSP*, 1992.

[5] M. Marcus, B. Santorini, and Marcinkiewicz. MA., "Building a large annotated corpus of english: The penn treebank," *Computational Linguistics*, vol. 19, 1993.

[6] J. Nivre, *Inductive Dependency Parsing*, Springer Verlag, 2006.

[7] S. Patwardan, S. Banerjee, and T. Pedersen, "Senserelate::targetword - a generalized framework for word sense disambiguation," in *Proc. of the Demonstration and Interactive Poster Session of the 43rd Annual Meeting of ACL*, Arbor, Michigan, 2005.

[8] M. Gugan and N. Hernandez, "Recognizing textual parallelism with edit distance and similarity degree," in *11th Conference of EACL*, Trento, Italy, 2006.

[9] D. Lin, "An information-theoretic definition of similarity," in *Proc. of the International Conference on Machine Learning*, Madison, Wisconsin, 1998.

[10] T. Pedersen, S. Patwardhan, and J. Michelizzi, "Wordnet::similarity - measuring the relatedness of concepts," in *Proc. of NAACL*, Boston, MA, 2005.

[11] T. Joachims, *Learning to Classify Text Using Support Vector Machines*, Kluver, 2002.