Prosodic Disambiguation in Spoken Systems Output

Samer Al Moubayed
KTH Centre for Speech Technology, Stockholm, Sweden
sameram@kth.se

Abstract

This paper presents work on using prosody in the output of spoken dialogue systems to resolve possible structural ambiguity of output utterances. An algorithm is proposed to discover ambiguous parses of an utterance and to add prosodic disambiguation events to deliver the intended structure. By conducting a pilot experiment, the automatic prosodic grouping applied to ambiguous sentences shows the ability to deliver the intended interpretation of the sentences.

1 Introduction

In using natural language in human computer interfaces, we expose ourselves to the risk of producing ambiguity – a property of natural language that distinguishes it from artificial languages. We may divide linguistic ambiguity broadly into lexical ambiguity involving single linguistic units and structural ambiguity – when an utterance can be parsed in more than one way as in:

"*I ate the chocolate on the desk.*" (1)

In many cases, structurally ambiguous utterances are not communicatively ambiguous as in:

"*I drank the water from the bottle*" (2)

The sentence in (2) has the same syntactic structure as in (1) but is not communicatively ambiguous as common knowledge resolves the ambiguity. In some cases, the structural ambiguity can lead to communicative ambiguity that needs to be resolved.

A growing body of research demonstrates that listeners are sensitive to prosodic information in the comprehension of spoken sentences. Rowles & Huang (1992) show how prosody can aid the syntactic parsing of spoken English in automatic speech recognition systems. Others have also associated pitch with prosodic grouping and disambiguation (e.g. Schafer et al., 2000), as well as pauses (e.g. Kahn et al., 2005). Allbritton, McKoon & Ratcliff (1996) conclude that speakers do not *always* use prosody to resolve ambiguity simply due to unawareness of its existence. There is also a great body of work on the use of prosody in computer generated speech, but to our knowledge there is no study to date on using prosody as a disambiguation tool in computer generated speech.

In this paper, we explore the possibility of automating prosodic disambiguation of computer generated speech in spoken dialogue systems to avoid communicating ambiguity. We assume that the system has access to the syntactic structure of the utterances it generates.

2 Placement of prosodic disambiguation

For the present purposes, we will assume a system modeling its possible utterances with binary CFG grammars, noting that any CFG grammar can be transformed into a binary one. A miniature grammar is provided in Figure 1, which generates a simple PP-attachment ambiguity. If a system produces such a potential communicative ambiguity, we need to know exactly where the ambiguity takes place in order to group the relevant sequence of words more clearly and prevent unintended interpretations. Figure 2 shows the parsing of the sentence: *"I ate <the chocolate on the desk>"* generated by the grammar in Figure 1, using chart parser style representation. In the chart, black trajectories are rules shared by all parses, green ones exist in the required parse tree only, and red ones are not part of the required parse trajectories while they exist in other parses. We see that the green trajectory must be grouped, as the words covered by this trajectory could otherwise be grouped in other ways, according to the grammar rules. Grouping them along the green trajectory distinguishes the intended parse from other parses. The trajectory is defined by its start and end nodes, hence the green trajectory is unique in that it is the only one starting and ending at those nodes.

R1: S → NP VP
R2: VP → V
R3: NP → Noun | Noun PP
R4: Noun → N | Det N
R5: PP → Prep NP
R6: V → Verb NP
R7: VP → V PP

Figure 1: Simple CFG grammar for English. The grammar generates PP-attachment ambiguity.

The disambiguation strategy suggested here, then, is to prosodically group a set of words only when not grouping them could result in a different parse, and ultimately a different interpretation.

3 A Pilot Experiment

As a listening test of the interpretation enhancement of the automatic disambiguation grouping of the previous algorithm, 15 sentences with coordination or PP-attachment ambiguities were generated using an in-house TTS. This system has a phrasing property implemented. 5 sentences of these were communicatively unambiguous but structurally ambiguous, and the rest were communicatively ambiguous. To ensure that the preferred meaning of these sentences is not taken into account, one subject had listened to these computer generated sentences without any grouping and gave her interpretation, we will call this subject "Subject A". Subsequently these sentences were introduced to two subjects after disambiguating them using prosodic grouping. These sentences contained PP-attachment and coordination ambiguity, and generated only two possible interpretations.

The results of these two subjects are grouped into two groups. The first one is the result of these subjects for sentences disambiguated to deliver the interpretation of the sentences which matched the one given by "Subject A", that is when the sentences do not receive any disambiguation. The other group is the results for the sentences delivering the opposite interpretation.

The result shows that 95% of the sentences received the correct interpretation after disambiguation when the desired interpretation matched this of "Subject A", while 75% of the sentences received the correct interpretation when the sentences disambiguated to deliver the other interpretation than "subject A" interpretation. In addition, the results show that the grouping using the proposed algorithm, as hoped for, did not affect the interpretation of the communicatively unambiguous sentences regardless of the prosodic disambiguation.

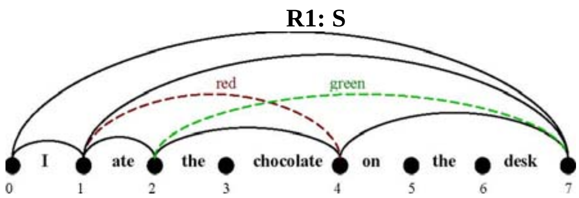

Figure 2: Possible parses of an example sentence. The black arcs are shared by all possible parses of the sentence. The green arcs exist only in the required parse and the red ones do not exist in the required parse by in other possible parses.

4 Conclusions

In this work, we presented an algorithm for spotting ambiguity in synthesized sentences with known syntactic structure. By conducting a small experiment, prosodic grouping (phrasing) is used by the disambiguation algorithm, and the results show high recognition rate by the subjects of the required interpretation of the disambiguation algorithm.

Future studies should focus on testing prosodic disambiguation using large scale grammar, or other types of grammars like PCFG, when disambiguation takes place depending on the probabilities of the multiple parses of the same utterance.

Acknowledgments

This work is supported by MonAMI, an Integrated Project under the European Commission's Sixth Framework Program (IP-035147).

Many Thanks to Jens Edlund for the interesting and helpful discussions.

References

Allbritton, D.W., McKoon, G., Ratcliff, R. "Reliability of prosodic cues for resolving syntactic ambiguity. Journal of Experimental Psychology: Learning, Memory, & Cognition, 22, 714-135, 1996.

Chris, Rowles., Xiuming, Huang. "Prosodic Aid to Syntactic and Semantic Analysis of Spoken English", in Proceedings of the 30st Annual Meeting of the Association for Computational Linguistic pages 112-119.ACL, 1992.

Jermy G. Kahn, Matthew Lease, Eugene Charniak, Mark Johnson, Mari Osterdorf. "Effective use of prosody in parsing conversational speech", Proceedings of the Conference in Human Language Technology and Empirical Methods in Natural Language Processing, p.233-240, Vancouver, British Columbia, Canada. 2005.

Schafer, Amy J., Shari R. Speer, Paul Warren & S. David White. "Intonational Disambiguation in sentence production and comprehension". Journal of Psycholinguistic Research, 29, 169-182. 2000.