

Non-traditional Prosodic Features for Automated Phrase-Break Prediction

Brierley, Claire

cb5@bolton.ac.uk
University of Bolton, UK

Atwell, Eric

eric@comp.leeds.ac.uk
University of Leeds, UK

The goal of automatic phrase break prediction is to emulate human performance in terms of naturalness and intelligibility when assigning prosodic-syntactic boundaries to input text. Techniques can be deterministic or probabilistic; in either case, the problem is treated as a classification task and outputs from the model are evaluated against 'gold standard' phrase break annotations in the reference dataset or corpus. These annotations may represent intentions (of the speaker or writer) or perceptions (of the listener or reader) about alternating chunks and boundaries in the speech stream or in text, where the chunking bears some relationship to syntactic phrase structure but is thought to be simpler, shallower and flatter.

In this paper, we begin by reviewing methodologies and feature sets used in phrase break prediction. For example, a tried and tested *rule-based* method is to employ some form of 'chink-chunk' algorithm (Lieberman and Church, 1992) which inserts a boundary after punctuation and whenever the input string matches the sequence: open-class or content word (chunk) immediately followed by closed-class or function word (chink), based on the principle that chinks initiate new prosodic phrases.

We discuss the limitations of using traditional features in the form of syntactic and text-based cues as boundary correlates, with illustrative experimental predictions from a shallow parser and evidence from the corpus. We then discuss the limitations of evaluating any phrase break model against a "gold standard" which itself only

represents one phrasing variant for an utterance or text.

There is an emerging trend of leveraging real-world knowledge to improve performance in machine learning, including speech and language applications. Nevertheless, we have diagnosed a deficiency of *a priori* knowledge of *prosody* in the feature sets used for the phrase break prediction task. In contrast, a competent human reader is able to project holistic linguistic insights, including projected prosody, onto text and to treat them as part of the input (Fodor, 2002). In this respect, multiple prosodic annotation tiers in the Aix-MARSEC corpus (Auran *et al.*, 2004) have been revelatory, since they capture the prosody implicit in text and currently absent in learning paradigms for phrase break models.

Insights such as: (i) the *transferability* of the chinks and chunks rule; plus (ii) the possibility of encoding a variety of prosodic phenomena (including rhythm and beats) in categorical labels (*cf.* the Aix-MARSEC corpus); plus (iii) an appreciation of prosodic variance gleaned from corpus evidence of alternative parsing and phrasing strategies, have informed the creation of ProPOSEL (Brierley and Atwell, 2008a; 2008b), a domain-independent prosodic annotation tool.

ProPOSEL is a **prosody** and **part-of-speech** English lexicon of 104,049 entry groups, which merges information from several widely-used lexical resources for corpus-based research in speech synthesis and speech recognition. Its record structure supplements word-form entries with syntactic annotations from four rival POS-tagging schemes, mapped to fields for: default open and closed-class word categories; syllable counts; two different phonetic transcription schemes; and lexical stress patterns, namely abstract representations of rhythmic structure (as in *201* for *disappear*, with secondary stress on the first syllable and primary stress on the final syllable).

We then contend that native English speakers may use certain sound patterns as *linguistic signs* for phrase breaks, having observed these same patterns at rhythmic junctures in poetry. We also contend that such signs can be extracted from canonical forms in the lexicon and presented as input features for the phrase

break classifier in the same way that real-world knowledge of syntax is represented in POS tags; and that like content-function word status or punctuation, such features are domain-independent and can be projected onto *any* corpus. One such sound pattern is the subset of complex vowels, which we define as the eight diphthongs, plus the triphthongs, of Received Pronunciation (Roach, 2000: 21-24).

Finally, we test the correlation between pre-boundary lexical items bearing complex vowels and gold-standard phrase break annotations on different kinds of speech via the chi-squared statistic, to determine whether the perceived association is statistically significant or not. Our findings indicate that this correlation is extremely statistically significant: it is present in contemporary, formal, British English speech (Brierley and Atwell, 2009) and seventeenth century English verse (Brierley and Atwell, 2010a); and it holds for spontaneous as well as read speech, and for multiple speakers (Brierley and Atwell, 2010b). We hypothesise that while complex vowels seem to constitute phrase break *signifiers* in English, this may translate to a subset of the vowel system in other languages.

References

- Auran, C., Bouzon, C. and Hirst, D.** (2004). 'The Aix-MARSEC Project: an Evolutive Database of Spoken British English'. *Proc. Speech Prosody*. 2004, pp. 561-564.
- Brierley, C. and Atwell, E.** (2008a). 'ProPOSEL: A Prosody and POS English Lexicon for Language Engineering'. *Proc. 6th Language Resources and Evaluation Conference*. LREC, 2008.
- Brierley, C. and Atwell, E.** (2008b). 'A Human-oriented Prosody and PoS English Lexicon for Machine Learning and NLP'. In *Proc. 22nd International Conference on Computational Linguistics*. Coling, 2008.
- Brierley, C. and Atwell, E.** (2009). 'Exploring Complex Vowels as Phrase Break Correlates in a Corpus of English Speech with ProPOSEL, a Prosody and PoS English Lexicon'. *Proc. INTERSPEECH'09*.
- Brierley, C. and Atwell, E.** (2010a). 'Holy Smoke: Vocalic Precursors of Phrase Breaks in Milton's Paradise Lost'. *Literary and Linguistic Computing*. **25(2)**.
- Brierley, C. and Atwell, E.** (2010b). 'Complex Vowels as Phrase Break Correlates in a Multi-Speaker Corpus of Spontaneous English Speech'. *Proc. Speech Prosody, 2010* (Forthcoming).
- Fodor, J. D.** (2002). 'Psycholinguistics Cannot Escape Prosody'. *Proc. Speech Prosody*. 2002, pp. 83-90.
- Lieberman, M. Y. and Church, K. W.** (1992). 'Text Analysis and Word Pronunciation in Text-to-Speech Synthesis'. *Advances in Speech Signal Processing*. Furui, S. and Sondhi, M. M. (ed.). New York: Marcel Dekker, Inc..
- Roach, P.** (2000). *Phonetics and Phonology: A Practical Course*. Cambridge: Cambridge University Press, 3rd Edition.