Why do we Accent Words?
The Processing of Information Structure and Prosodic Structure

Sasha Calhoun
School of Philosophy, Psychology and Language Sciences
University of Edinburgh
Sasha.Calhoun@ed.ac.uk

One of the main functions of prosody is to signal *information structure*: i.e. to "chunk" the information in a spoken utterance into coherent units; and to "highlight" the important, or *focussed* (as per Rooth 1992), element(s) in each unit (e.g. Halliday 1968). However, as much work has shown, it is problematic to directly equate prosodic phrases with information units, and pitch accents with semantic focus (e.g. Ladd 1996, Truckenbrodt 1995, Wagner 2006). I propose that, rather than a direct relationship; information structure acts as a strong probabilistic constraint on the alignment of words with metrical prosodic structure (see also Calhoun 2006). Foci try to align with nuclear accents; however realisation is also affected by other factors, including lexical predictability and syntactic structure, constraints on prosodic structure itself, as well as pragmatic distinctions. In support of this model, I present an analysis of accent prediction results using a substantial corpus of annotated spontaneous speech. Finally, I seek to motivate the model in terms of language processing. Why should we compute a separate prosodic structure rather than just using the phonetic cues to prominence and phrasing directly? Does the "highlighting" function of prosody really separate into both semantic focus and lexical predictability? Why?

Experimental evidence comes from accent prediction models developed using the NXT Switchboard corpus (Calhoun et al. 2005). A subset of 8915 words from 18 conversations in the corpus were annotated for prosodic features, including nuclear and non-nuclear accents and phrase breaks; focus status; and lexical and syntactic features including part-of-speech, syntactic clauses and accent ratio (a measure of a word's "accentability" shown to perform better than other standard lexical predictability measures, Nenkova et al. 2007). Multiple regression models were trained to predict the accent status (nuclear, non-nuclear or unaccented) of each word, using different feature sets, i.e. phrasal (e.g. position in the phrase, words since last accent), semantic/syntactic (e.g. focus status, position in clause, accent ratio) and combined (phrasal + semantic/syntactic). The phrasal feature model performed better than the semantic/syntactic model and only slightly worse than the combined model. This is consistent with the view that accents are manifestations of metrical structure, rather than appearing independently to mark semantic or lexical features. Crucially, both the semantic/syntactic and combined models could reliably predict the occurrence of nuclear accents, but not non-nuclear accents. Focus status was more strongly predictive of nuclear than non-nuclear accents, and vice versa for accent ratio. These results suggest that focus acts mainly as a constraint on the position of the nuclear accent, while the appearance of non-nuclear accents are determined by other factors, such as the inherent accentability of the word and rhythm.

My model and accent prediction results offer support for a coherent, metrical prominence structure; rather than direct use of phonetic cues to prominence in processing. It has been

proposed that prosodic phrasing acts to "unify" different levels of linguistic representation of an utterance (e.g. Frazier et al. 2006). I suggest this is also achieved through the hierarchical patterns of metrical prominence. The highly rhythmical, and predictable, nature of these patterns may facilitate this. Further, these results seem to show that the "highlighting" function of prosodic prominence needs to be separated into the marking of semantic focus and lexical predictability, as these have distinct effects on accent realisation. I suggest this is because the two operate at quite different levels of language processing. The signalling of semantic focus, which affects truth-conditional and pragmatic interpretation, is part of the high-level planning of an utterance; concurrently with the overall planning of prosodic structure. Lexical predictability, on the other hand, seems strongly related to lexical retrieval; and therefore principally affects low-level within-phrase prominence patterns. It is hoped that this discussion, along with my theoretical model and experimental results, will help in the move towards more sophisticated use of both prosody and information structure in models of language processing.

References

Calhoun, S. 2006. Information Structure and the Prosodic Structure of English: a Probabilistic Relationship, PhD thesis, University of Edinburgh.

Calhoun, S., M. Nissim, M. Steedman & J. Brenier 2005. A Framework for Annotating Information Structure in Discourse. In 'Frontiers in Corpus Annotation II', *ACL Workshop*, Ann Arbor, MI.

Frazier, L., K. Carlson & C. Clifton Jr. 2006. Prosodic phrasing is central to language comprehension, *Trends in Cognitive Science* 10, 6, 244-249.

Halliday, M. 1968. Notes on transitivity and theme in English: Part 3, *Journal of Linguistics* 4, 179–215.

Ladd, D. R. 1996. *Intonational Phonology*, Cambridge University Press, Cambridge, UK.

Nenkova, A., J. Brenier, A. Kothari, S. Calhoun, L. Whitton, D. Beaver & D. Jurafsky 2007. To memorize or to predict: Prominence labelling in conversational speech. In *HLT:NAACL*, Rochester, NY.

Rooth, M. 1992. A theory of focus intepretation, *Natural Language Semantics* 1, 75–116.

Truckenbrodt, H. 1995. Phonological Phrases: Their Relation to Syntax, Focus and Prominence, PhD thesis, MIT.

Wagner, M. 2006, Givenness and locality, in M. Gibson & J. Howell, eds, '*Proceedings of SALT XVI*', CLC Publications, Ithaca, NY.