Sentence prosody leaks into the lexicon: evidence from Mandarin Chinese

Synopsis: While the precise extent to which phrasal phonology interacts with word-level phonology is a long-standing issue, it is generally assumed that lexical phonology is at least somewhat independent of phrasal phonology, including intonation. Exemplar theory complicates this division, as phonetically detailed exemplars encode context-dependent prosody in the lexical representation (Pierrehumbert 2016). In line with this prediction, some evidence for the lexical encoding of intonation has been found in German and English, languages in which pitch accents are assigned at the phrasal level (Schweitzer et al. 2015). Schweitzer et al. showed that f0 contours are more stable in predictable collocations than in unpredictable collocations, suggesting a possible lexicalization of intonation. The current study probes this issue in Mandarin Chinese, a language with lexical tone and a well-studied system of prosodic focus marking. We ask whether sentence level prosody (focus) can be similarly lexicalized in a language that already has lexical tone.

Approach: Words that receive prosodic focus in Mandarin have higher pitch, longer duration, and greater intensity (e.g., Wang and Xu 2011). Notably, these same phonetic dimensions (pitch, duration, intensity) also distinguish Mandarin lexical tones (e.g., Whalen and Xu 1992). If focus prosody is lexicalized, then words that tend to be focused more often will show higher pitch, longer duration and greater intensity (even when not focused). As a proxy for how likely a word is to receive focus, we calculated the average predictability, also known as the "informativity", of each word. The key assumption in approximating focus with informativity is that unpredictable words are likely to constitute new information and therefore receive focus. Informativity is defined as the negative log average contextual predictability of a word in every context in which it appears in, weighted by the contextual predictability of the contexts (Piantadosi et al. 2011). The numerical expression is given in equation 2, where *c* is a context, *C* is the set of all contexts, *w* is a word and *W* is the set of all words.

Equation 2: $-\sum_{c} Pr(C = c \mid W = w)\, log_2\, Pr(W = w \mid C = c)$

Work on English has shown that the duration of words is related to informativity, even after word frequency and local predictability are controlled for (Seyfarth 2014). That is, a word that is predictable on average (high informativity) tends to be short even when it is in a locally unpredictable context. This suggests that the phonetic effects of predictability are lexicalized to some extent. We hypothesize that focus prosody is the force behind predictability effects on word duration (see also Aylett and Turk 2004). On this account, the reason why informativity conditions variation in duration is because unpredictable words are more likely to receive focus, as "new information". Our interpretation of this past work leads us to our **main hypothesis**: phonetic indicators of focus will vary with informativity, such that high informativity words take on the phonetic characteristics of focus position.

Method: To test our hypothesis, we calculated word informativity in Mandarin Chinese using frequency statistics estimated over a 431 million word subtitle corpus. The acoustic information of content words was then extracted from a large telephone conversation speech corpus with over 400,000 tokens and 6,000 word types spoken by

1,655 individuals. Linear mixed effects models were fit to measures of word duration, maximum intensity and maximum f0, three phonetic parameters related to focus position in Mandarin. The model structure is given in (1). Random effects included intercepts for item, tone sequence, and speaker, and a by-speaker random slope for informativity. The key fixed effect was informativity, which was included along with 15 other control variables, including frequency, local predictability and speech rate.

Main result: Informativity significantly influenced all three variables known to be associated with focus in Mandarin Chinese. A unit increase in informativity conditioned higher maximum pitch (β: 0.0039, SE: 0.0008, *t:* 4.6575), longer duration (β: 0.0099, SE: 0.0014, *t:* 7.0700) and higher intensity (β: 0.4043, SE: 0.0696, *t:* 5.8113). The positive coefficients indicate that the directions of the effects are in the direction predicted by the lexicalization of focus.

Discussion: Learning independent (or quasi-independent) phonological generalizations at the lexical and phrasal levels of grammar entails that listeners can parse layered influences on phonetic form into their respective phonological sources. Our results show that intonational context can leave a "prosodic residue" in the lexicon. Even after controlling for numerous factors, including word frequency and local predictability, our proxy for focus likelihood (informativity) explained significant variation in pitch, duration and intensity across words of Mandarin Chinese. Possibly, listener partitioning of phrasal and lexical influences on phonetic form is imperfect. Listeners fail to attribute the entire influence of focus to prosodic context, c.f., incomplete compensation for coarticulation, retaining some residual influence of focus in the lexical representation. Imperfect attribution is perhaps amplified in this case by the fact that lexical tone contrasts in Mandarin make use of the same phonetic dimensions as focus. Our account requires phonetically detailed lexical representations alongside grammatical mechanisms for abstracting phonological generalizations at the lexical and phrasal levels (along the lines of Pierrehumbert 2016).

(1) Model structure: *Dependent variable (either Duration, Maximum intensity, Maximum pitch) ~ Frequency + Forward predictability + Backward predictability + Forward informativity + Backward informativity + Word length + Preceding disfluency + Following disfluency + Preceding pause duration + Following pause duration + Preceding speech rate + Following speech rate + Previous self-mention + Previous cross-speaker mention + Age + Gender + Syntactic category + (1 | Word type) + (1 | Tone sequence) + (1 + Forward informativity + Backward informativity | Speaker)*

References

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1), 31-56; **Piantadosi, S. T., Tily, H., & Gibson, E.** (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108(9), 3526-3529; **Pierrehumbert, J. B.** (2016). Phonological representation: beyond abstract versus episodic. Annual Review of Linguistics, 2, 33-52; **Schweitzer, K., Walsh, M., Calhoun, S., Schütze, H., Möbius, B., Schweitzer, A., & Dogil, G.** (2015). Exploring the relationship between intonation and the lexicon: Evidence for lexicalised storage of intonation. *Speech Communication*, 66, 65-81; **Seyfarth, S.** (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, 133(1), 140-155; **Wang, B., & Xu, Y.** (2011). Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese. *Journal of Phonetics*, 39(4), 595-611.