Phonology Constrains the Distribution of the Particle *lah* in Singapore English

James Sneed German[1] *& Laurent Prévot*[2]

[1]Nanyang Technological University; [2]Aix-Marseille Université & the Laboratoire Parole et Langage

jsgerman@ntu.edu.sg, laurent.prevot@lpl-aix.fr

The meaning of the particle *lah* in Singapore English (SgE) has been characterized as serving a variety of social and discourse functions (see Besemeres & Wierzbicka (2003) for a review). All of these functions, however, reflect a choice on the part of the speaker to convey something. At the same time, *lah* has specific positional and prosodic properties which affect its realization. For example, it occurs only at the ends of utterances and before major prosodic breaks. As *lah* occurs primarily in statements and directives, its realization is governed by the patterns that normally occur in those utterances, such as the presence of a final f0 fall. While some accounts assume distinct tonal variants for *lah* (Wong, 2004), it is never prominence-bearing, and the most common realization involves a falling or sustained low f0 contour.

Utterances differ in terms of the phonological properties of the lexical items that they end with. In particular, they may end with shorter or longer sequences of unstressed syllables ("*He wants to volun**teer**" vs.* "*He's doing exercises*"). Since *lah* is unstressed, adding it to the end of an utterance will effectively increase the length of the unstressed sequence. According to principles of eurhythmy (e.g., Hayes, 1984), such long unstressed sequences should be dispreferred. This raises the possibility that constraints on metrical structure may limit the extent to which a speaker can freely deploy *lah* for pragmatic or social functions. In other cases, the phonological context may actually *favor* the presence of *lah*. When the final syllable is stressed, for example, the inclusion of *lah* can relieve tonal crowding by supplying additional segmental material over which to realize the extra tones. If *lah* is sufficiently consistent with the pragmatic context, then the speaker might opt to use it when he or she might not have otherwise.

In sum, there is reason to suppose that the distribution of *lah* is partly determined by the local prosodic context. We tested this hypothesis through analysis of a text-based corpus of conversational spoken Singapore English (ICE-Singapore). We restricted our analysis to the "Private Dialogues (S1A)" in which *lah* is well-represented (1,586 out of 213,555 word tokens). We extracted the last five syllables of each of the 29,855 utterances, of which 942 ended with *lah*. Excluding words not in our lexical database, this yielded 892 contexts with *lah* and 26,952 without. Overall SgE preserves the stress pattern of British English (Bao, 2006), so utterance-final stress patterns were estimated by cross-indexing wordforms with the Celex2 lexical database (Baayen, et al., 1996). As shown in Table 1(left), we categorized tokens according to the number of unstressed syllables preceding the target location (i.e., the utterance boundary or *lah*).

Table 1. *Left*: Frequency of utterance-final contexts with and without *lah* by stress pattern. *Right*: Percentage deviation from the expected value under the null hypothesis.

Stress	No *lah*	*lah*
xxx1	17969	644
xx10	7055	198
x100	1682	44
1000	219	6
0000	27	0

Stress	No *lah*	*lah*
xxx1	-0.3%	+7.9%
xx10	+0.5%	-14.9%
x100	+0.7%	-20.5%
1000	+0.6%	-16.8%
0000	--	--

- "0" : unstressed syllable
- "1": stressed syllable
- "x": syllable of any type or utterance boundary

If the distribution of *lah* is not sensitive to the local prosodic context, then its distribution across categories should reflect the overall proportions across categories. A Chi-square test revealed that this distribution deviates significantly from the expected one ($\chi^2 = 11.77$, $p < 0.01$). ("0000" includes no *lah* tokens and was therefore excluded from the analysis.) Table 1(right) gives the percentage deviation for each cell, showing that the effect is in the expected direction. In other words, *lah* is underrepresented in contexts where its inclusion would have resulted in an unstressed string of two or more syllables, and overrepresented when the final syllable is stressed. These findings support the view that the distribution of *lah* is partly determined by the metrical structure of the local lexical context, and is not solely a matter of speaker intention. These are closely related to findings by Calhoun (2010) and German et al. (2006) showing that phonological constraints can influence accent placement independently of discourse pragmatic considerations. Our findings differ in that *lah* represents a choice of segmental, rather than purely intonational structure. The general pattern is predicted to extend to other particles of SgE as well as similar particles in other languages (e.g., French *quoi*).

References

Baayen, R.H., Piepenbrock, R., & Gulikers, L. (1996). Celex2. *LDC96L14, Linguistic Data Consortium, Philadelphia.* / Bao, Z. (2006). Clash avoidance and metrical opacity in Singapore English. *STUF-Sprachtypologie und Universalienforschung*, *59*(2/2006), 133-147. / Besemeres, M., & Wierzbicka, A. (2003). Pragmatics and cognition: The meaning of the particle *lah* in Singapore English. *Pragmatics & Cognition*, *11*(1), 3-38. / Calhoun, S. (2010). How does informativeness affect prosodic prominence? *Language and Cognitive Processes*, *25*(7-9), 1099-1140. / German, J., Pierrehumbert, J. & Kaufmann, S. (2006). Phonological constraints on nuclear accent placement. *Language 82(1)*, 151-168. / Hayes, B. (1984). The phonology of rhythm in English. *Linguistic Inquiry*, *15*(1), 33-74. / Wong, J. (2004). The particles of Singapore English: A semantic and cultural interpretation. *Journal of Pragmatics*, *36*(4), 739-793.