Language-specificity in Speakers’ Strategies of Focus Expression

Martin Ho Kwan Ip, Anne Cutler; The MARCS Institute; Contact email: m.ip@westernsydney.edu.au

Information structure is a cognitive universal (because our experience mixes the familiar and the unexpected) and its expression in speech is a linguistic universal (because we talk about our experience). However, this is not to say that the way information structure is realized by prosody is always the same. It varies across languages, in that speakers of some languages do not call on prosodic resources to express focus when rich morphosyntactic means can do the same job (e.g., in Wolof [1]). And even where prosody is the way of choice for marking focus, its acoustic components (e.g., duration, pitch) may be exploited differently across languages [2]. Prosodic strategies for expressing information structure can also vary across individuals [3] and discourse contexts [4].

The present report draws on an unusually extensive collection of focus production data in which 24 native speakers of each of Mandarin Chinese and English were recorded engaging in structured dialogues. All participants engaged in two structured dialogues in their native language, where each dialogue contained pairs involving the same word tokens in a focused vs. an unfocused realization. In both languages, the dialogues differed in context: the first dialogue involved a buyer negotiating with a street vendor, while in the second dialogue a high-school student and a police inspector discussed a past event. The Mandarin and English texts were comparable (translations), and 21 pairs of focused and unfocused words were collected for each language (19 out of the 21 word pairs in each language were either monosyllabic or disyllabic). Dialogue construction was guided by Krak’s [5] proposal of the different pragmatic contexts that elicit focus (e.g., correction, response to wh-questions); unfocused tokens were given information. For each of the word pairs, we measured 6 parameters of prosodic productions: duration, mean F₀, maximum F₀, F₀ range, mean rms-intensity, maximum rms-intensity. This produced a total of 12,096 measurements (2 languages x 21 pairs x 2 focus levels x 6 prosodic parameters x 24 participants), sufficient to support examination at an individual level.

Figure 1 shows the results for inter-speaker variability in English and Mandarin productions of prosodic focus. For each prosodic parameter, difference scores were calculated to measure the relative amount of increase in prosody from the unfocused to the focused tokens. In Mandarin, analyses for the first dialogue revealed a significant negative correlation between use of F₀ range and maximum intensity (r = -.461, p = .023). Similarly, all other F₀ measures were negatively associated with maximum intensity, although not significantly. Further, participants who increased their maximum intensity to signal focus would also tend to produce greater duration (r = .517, p = .010). This indicates that Mandarin speakers differ in their preferred prosodic strategies to mark focus, based on whether they would produce greater F₀ range or maximum intensity and duration. The English data show a different pattern of preference, however, where this time, speakers’ increase in maximum intensity was significantly associated with an increase in F₀ (mean) (r = .650, p < .001).

Another question that we addressed was whether participants were consistent in their use of the different prosodic parameters across all instances of focus expression (see Figure 2). Mandarin speakers showed more intra-speaker variability in their increase in duration (SD = .060) compared to English speakers (SD = .045), t(1, 46) = 2.33, p = .024. Further, Mandarin speakers also showed more variability in their F₀ range (SD = 50.55) than English (SD = 33.56), t(1, 46) = 3.47, p = .001. On the other hand, English speakers showed greater variability in their use of mean F₀ across the focused tokens (SD = 32.14) than Mandarin speakers (SD = 19.19), t(1, 46) = 3.53, p = .001.

Results from the second dialogue showed different patterns of inter- and intra-speaker variability. For inter-speaker variability, increase in mean F₀ was associated with increase mean intensity in both English (r = .433, p = .035) and Mandarin (r = .427, p = .037). Mandarin participants also showed a positive association between mean F₀ and maximum intensity (r = .553, p = .005). Furthermore, as in the first dialogue, Mandarin speakers who increased their duration for focus were also more likely to increase their maximum intensity (r = .477, p = .019). For variability across each speaker, only duration showed a significant effect, in which Mandarin speakers (SD = .068) showed greater intra-speaker variability than English speakers (SD = .041), t(1, 46) = 5.52, p = .015.

The dialogues used in this study were semantically controlled and we further attempted to control the vowels used in the two sets of word tokens on height and backness as far as possible. Each of the two speaker groups was largely female (21 and 22 for English and Mandarin respectively). Nonetheless we see significant differences across the two language groups in how individual speakers
vary in their exploitation of the different parameters. Since Mandarin and English, despite the differences in their respective phonology, are highly alike in the parameters (duration, pitch range, intensity) used to effect focus [6, 7], the present observed difference is particularly interesting, as it may reflect language-specific differences in listeners’ perception of prosodic focus. For example, Mandarin speakers’ preferences for either F0 (range) or maximum intensity in the first dialogue suggest that Mandarin listeners would accept either increased F0 or increased intensity as focus, while the lack of discrimination in English speakers suggests that English listeners only recognize focus when increased F0 is coupled with higher intensity. At the same time, it is intriguing that discourse contexts could affect production of focus. For example, unlike the first dialogue, focus in the second dialogue was realised in the same way in both English and Mandarin (i.e., an increase in intensity and F0 for focus for both languages). This may be due to the fact that the first and second dialogue provided a different case of focus, where the former involved negotiation of a purchase and the latter involved a police interrogation. Follow-up analyses should look at grammatical/structural differences in the two dialogues; the first dialogue involving a conversation with a street vendor may be written in a more colloquial style compared to the second dialogue. Therefore, the encoding of information structure may be universal, and the range of parameters in which it is encoded may also be universal, but individual differences in the selection from this range exist, and, most interestingly, the options open to individuals here may differ across contexts and languages.

**Figure 1.** Significant inter-speaker variability in the first dialogue (above) and the second dialogue (below) across Mandarin (red) and English (blue).

**Figure 2.** Significant language-specific differences in intra-speaker variability in the first dialogue (first three panels) and the second dialogue (far right panel) in Mandarin (black) and English (grey).

**References**