Information structure and prosodic prominence: how does sentence final particle affect Cantonese intonation?

Hong Zhang¹, Mark Liberman¹, Tan Lee²

¹University of Pennsylvania, USA
²Chinese University of Hong Kong, Hong Kong, China
zhangho@sas.upenn.edu, myl@babel.ling.upenn.edu, tanlee@cuhk.edu.hk

Abstract

This study addresses the question of whether and how morphosyntactic cues to information structure may affect sentence prosody in Cantonese. Cantonese Sentence Final Particles (SFP) are considered as the optional marker for information structure in the colloquial form of the language. A Cantonese version of the Map Task Corpus has been collected. Regression analyses have found a trade-off relation between prosodic and morphosyntactic means for focus marking. However, the effect size is rather small, and may not be auditorily significant. This study also illustrates a method for efficient collection of speech data for future research.

Index Terms: Cantonese, sentence final particle, information structure, prosodic prominence

1. Introduction

A vast body of research has suggested that in addition to syntactic structure and focus sensitive lexical items, prosody also appears to be crucial in signaling the information status of sentence elements [1, 2, 3, 4]. In addition to the widely studied focus-marking operators such as topicalization and embedding [5], Chinese languages can also associate morphosyntactic with information structure through the use of a class of lexical items called sentence final particles (SFP). Cantonese is especially rich in both the inventory and meanings conveyed through SFPs. In particular, the use of SFPs is often regarded as a prominent feature of colloquial Cantonese.

Although consensus has not been reached regarding the meaning of each individual SFP [6, 7, 8, 9, 10], it is relatively clear that the connection between SFPs and information structure can be realized in at least three ways. First, there has been discussion on the so-called focus SFPs [11, 12]. It is proposed that these SFPs behave in a manner analogous to quantifiers in both syntax and semantics. Thus attempts have been made to treat them as focus sensitive operators such as ‘only’ and ‘also’. The second possibility is through the pragmatic meanings inherent to SFPs. Particles that encode procedural meanings or indicate the relevance of certain components within a sentence [7, 13] may allude to the notion of new information or contrast under particular pragmatic contexts. The third possibility is realized via more intricate interaction between SFPs and syntactic movement [14]. For example, SFPs are an indispensable component in Dislocation Focus Construction, where the on-focus arguments are dislocated to the beginning of the sentence.

Given the apparent link between SFPs and information structure, and the widely assumed role of information structure in sentence prosody, Cantonese provides an interesting testing ground for a case study of the relation between information structure and prosody with and without mediation from morphosyntactic tools. The goal of the present study is therefore to investigate the mapping between prosodic realization and morphosyntactic marking through the use of SFPs of information structure in Cantonese.

2. Data collection

2.1. The Map Task design

Since the use of SFPs is highly colloquial, it is necessary to keep a balance between sufficient control of the information content of utterances and speech style in order for proper data generation. A version of the Map Task [15] design was adopted to meet this end.

The task was designed to be accomplished by coordination between two participants. Each map was drawn with a set of landmarks as reference points for speakers. The Instruction Giver’s map was presented with a designated route, while the Instruction Follower’s map was not. The goal for the pair of participants was to reconstruct the route presented on Instruction Giver’s map on Instruction Follower’s version. Figure 1 shows an example pair of maps adapted from the HCRC Map Task Corpus[15].

16 pairs of maps were borrowed and adapted from the HCRC Map Task Corpus. The shape and name of landmarks were checked by two native speakers of Cantonese to ensure that the landmarks were readily recognizable by and the names conform to the vocabulary of native speakers. The information structure of the speech production contexts was controlled by different aspects and degrees of sharedness of the landmarks. Six levels of sharedness were considered to encode shared information, new-old information and four types of contrasts: size,
number, location and name. The manipulations included: 1. the same design, name and location on each of the maps; 2. the presence and absence of a same object at the same location; 3. same object appears at the same location with different sizes; 4. same object appears at the same location with different quantity; 5. objects that appear twice at different locations on either version of the map; and 6. the same object appears at the same location with different names. In addition, a compass was printed on the background of maps that does not interfere with main landmarks to help with direction instructions. Each map contained about 12 to 15 landmarks, about 40% of which were mismatched in the manner described.

2.2. Experiment procedure

The entire data collection process was conducted through the online conferencing platform Zoom [16]. Participants were recruited both from Hong Kong and the Cantonese community in Philadelphia. Speakers were invited to join the experiment from a quiet location to their best convenience. Each speaker who signed-up for the experiment was encouraged to bring a friend to the meeting. Familiarity between interlocutors may affect the speech style of the conversation [17]. By controlling for the familiarity between experiment participants, it was hoped to encourage the speech style to be as colloquial as possible.

During the experiment, each speaker was sent a map at the beginning of each session of the task. The role of Instruction Giver and Instruction Follower alternated in each session, with the initial role assigned randomly. The Instruction Follower was required to reconstruct the route as accurate as possible, and send the reconstructed map back at the end of the session. The speakers were informed that the maps were not exactly the same in each session but they were not aware of where and how the maps might differ. When referring to particular objects on the map, speakers were instructed to first use the names that were marked on the map. Three to five sessions were completed by each pair of speakers.

Recordings were made with Zoom’s built-in recording function via computer mounted microphones on participants’ device and saved in separate channels on experimenter’s computer in M4A format. A total of 12 speakers (7 female, 5 male) participated in the data collection, contributing about 5 hours of speech.

3. Data processing and annotation

Original recordings were converted to WAV format with sampling frequency of 16kHz. On average, each speaker provided between 120 and 200 sentences that contain clear information structure reference. A sentence was defined as an utterance profiled to certain landmarks as the goal of particular movements. The arguments that have been identified as in the on-focus position of utterances included: the name of the landmark under reference; the relative position of landmarks, such as the designated end point of a movement requested by Instruction Giver: “to the right of the barn”, and the shape of the route, such as the phrase “a circle” in “make a circle at the upper left corner.

Sentences were categorized by whether ended with an SFP independent of their information structure. This is because some SFPs are mostly used to stress the noteworthiness of the proposition in proper conversational contexts, such as the particles wo3, maa3 and aa3. The determination of correspondence between particular SFPs and information structure became problematic in such scenarios. Among the annotated sentences, 1101 of them did not end with an SFP, which accounts for 57.4% of the corpus.

4. Results

4.1. The prosody of SFP

Although not the primary focus of the current study, the prosody of SFP itself may be indicative of the information structure of the sentence that worth mentioning. Two broad patterns of association between SFP and sentence body intonation contour can be identified from the data. The first is characterized as the continuation of sentence body contour, and the second is that the SFP itself forms a separate prosodic unit. Figure 2 lists example sentence contours from a male speaker. 2(a) and 2(b) are instances where the contours of SFP, laa1 and ge3, are continued from the body of respective sentences. 2(c) and 2(d) illustrate two instances where an SFP constitutes its own prosodic unit. Ge3 in 2(c) has a raised starting pitch level and longer duration, compared to the syllables in the sentence body. Laa3 in 2(d) has a rising contour, rather than its inherent mid-level tone, which apparently breaks from the preceding sentence contour.

The presence or absence of prosodic boundary of some kind may indicate the phrase an SFP attached to. 2(a) and 2(b) both have a focus on the penulement of the entire utterance, where the speaker was stressing the relative position with regard to certain landmarks as the goal of particular movements. 2(c) has a focus on the first word of the phrase preceding the SFP, where the quantity of the landmark (leun-ge, two) is contrasted. 2(d) is a little bit tricky to neatly fit into the two categories identified above. However, the brief pause before reference to the landmark (zi-naam-zam, the compass) indicates that the entire phrase zi naam zam ge sai bin (to the west of the compass) should be interpreted as the focus by the listener.

This crude observation of limited samples suggests possible connections between the intonation contour of SFP and its scope taking in sentence structure. Cues such as duration, contour shape and pitch level may indeed predict the information structure and syntactic phrasing involving particular SFPs. The exact relation among the pitch contour of SFP, information structure and syntactic/prosodic phrasing warrants a separate discussion.
The pitch contour of SFP is continuation from sentence body but with longer duration.

The pitch contour of SFP is the continuation from sentence body.

The SFP forms its own prosodic unit with apparent prosodic boundary.

The SFP forms its own prosodic unit by altering the pitch contour.

Figure 2: Example utterances with SFP.

4.2. Sentence prosody

Measurements were grouped independently by the presence of SFPs, type of information structure, and the location of on-focus argument of the sentence (sentence final vs. non-final). The position of on-focus arguments was considered because F0 declination over the duration of an utterance can result in a prominent argument appear in a prosodically weak position. Syntactically, the default position for on-focus argument is at the right edge of the most embedded clause [5], which is often prosodically weak.

F0 range was calculated by taking the 2.5% and 97.5% quantile of the total measurements in sentence or phrase to minimize the bias introduced by octave jump and other measurement errors. Median RMS amplitude is the median value of all the measurements in segments of interest.

4.2.1. Prosody at sentence level

Figure 3 shows density plots of F0 range and median RMS amplitude of sentences with and without SFP in two focus positions. 3(a) and (b) suggest that regardless of focus position, sentences with and without SFP have similar F0 range distribution. The median amplitude is a little smaller for sentences that do not end with an SFP. The difference is about 200 amplitude units for both focus positions.

Here we consider the effects of information structure, focus position, and SFP on F0 range and median RMS amplitude of sentences. Ordinary least square (OLS) regression is used to find the effect of each variable on two measurements of prosodic prominence. Two additional predictors, the duration of the sentence and the focus-to-background ratio, calculated by dividing the duration of on-focus phrase by sentence duration, are also included, as shorter sentences may be produced with greater energy, and higher focus-to-background ratio may have similar effect. Interaction terms of information structure and SFP, as well as sentence duration and on-focus ratio are also included in regression models.

Table 1 summarizes the regression result for F0 range. This model suggests that among the predictors of our primary interest, the presence of SFP in a sentence does have significant effect on the F0 range of the entire sentence, when all other things are controlled. In particular, compared to sentences without SFP, sentences with an SFP are expected to have an F0 range of 0.82 semitone higher on average.

Table 1: Summary of the regression model that finds the effect of information structure and SFP on F0 range at sentence level

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>Std error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence duration</td>
<td>2.163</td>
<td>0.339</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Log foc-ratio</td>
<td>-0.908</td>
<td>0.547</td>
<td>0.097</td>
</tr>
<tr>
<td>Info-structure type</td>
<td>0.534</td>
<td>0.405</td>
<td>0.187</td>
</tr>
<tr>
<td>Position: non-final</td>
<td>0.002</td>
<td>0.245</td>
<td>0.99</td>
</tr>
<tr>
<td>S-type: SFP</td>
<td>0.82</td>
<td>0.249</td>
<td>*0.001</td>
</tr>
<tr>
<td>Duration*ratio</td>
<td>0.631</td>
<td>0.18</td>
<td>*0.0005</td>
</tr>
</tbody>
</table>

The second regression model explores the effect of SFP and information structure on the median RMS amplitude of sentences. Same model specification is used as the first model described above. Table 2 summarizes the regression result of this model.

The fit of the model indicates that both information structure type and the position of on-focus argument have significant effect on the median RMS amplitude of the sentence. However, SFP does not turn out to be significant. Specifically, when every other variable is held constant, non-contrastive focus would be expected to raise the median amplitude by an average of 121.55 RMS units. This effect is significant at $p = 0.041$.

4.2.2. Prosody of the on-focus phrase

Figure 4 is the phrase version of figure 3, which shows the distribution of both F0 range and median RMS amplitude of the on-focus phrase. 3(a) and (b) suggest that sentences with SFP
5. Discussion

In this study, a corpus of spontaneous speech of Cantonese produced in semi-controlled pragmatic settings has been used to answer whether information structure and SFPs would affect sentence and phrase level focus prosody. Results from regression analyses found different effect on different acoustic correlates of prosodic prominence by SFPs and information structure.

The presence of SFP expands the F0 range of the entire sentence, while narrows down the F0 range of the on-focus arguments. This result provides evidence for an effect of SFP on sentence intonation, and the use of SFP may indicate the increase of sentence prominence in terms of wider F0 range. On the other hand, the narrowed F0 range for on-focus arguments in sentences with SFP implies a trade-off between prosodic prominence and additional morphosyntactic apparatus which achieves the same function. That is, with the presence of SFP, the necessity for indicating focus with some prosodic cue will be reduced. However, it should also be noted that these effects are relatively small, with a magnitude less than one semitone (which translates to a difference less than 5 Hz). Therefore this effect is more likely to just involve pre-planning during speech production but has little perceptual significance.

The regression models have also found an effect of focus position on phrase level F0 range, with non-final focus being less prominent, in terms of both F0 range and amplitude. This result shows that when the requirement for prominence contradicts the default prominent position decided by sentence structure, some prosodic cues may be used to compensate for the conflict. Thus when the on-focus arguments are in a prosodically prominent position (i.e., early in the sentence), little additional effort is needed to further mark the argument under focus. This difference is about 1.6 semitone, which can potentially be perceptually meaningful.

However, information structure is only able to find a marginally significant effect on sentence median amplitude. This finding may add a little piece of evidence in support for the existence of a prosodic difference between contrastive and non-contrastive focus [20].

6. Conclusions

This study attempts to address the question of whether and how Cantonese SFP and information structure may affect sentence and phrase prosody. The acoustic analyses suggested the possible existence of a trade-off between morphosyntactic and prosodic means in information structure marking, although the effect may not be perceptible. The method used for Cantonese Map Task collection also enables more efficient speech data acquisition for production studies.

7. Acknowledgements

This project is partly supported by a grant awarded to Prof. Tan Lee.
8. References


