THE INTERPLAY AND INTERACTION BETWEEN PROSODY AND SYNTAX: EVIDENCE FROM MANDARIN CHINESE

Chiu-yu TSENG and Da-de CHEN
Institute of Linguistics (Preparatory Office)
Academia Sinica
Taipei, China 115-21
e-mail: cytling@sinica.edu.tw

ABSTRACT

This paper aims to (1.) quantify possible correlation between syntactic structure and prosodic manifestation in Mandarin Chinese, (2.) explore to what extent such correlation could be predicted by syntactic structures and what may go beyond these correlations, and (3.) increase more operational characteristics to Mandarin prosodic structures.

INTRODUCTION

A major portion of the previous works by our group on a Mandarin Chinese speech database collected in Taiwan have been concentrated on developing tools to process the collected speech data [1]. Since we attempted to process the speech data from the physical perspective by employing a phonetic oriented approach, the development of these tools depended largely on how we characterized the major phonetic features derived. We also aimed to characterize the major phonetic as well as prosodic properties of Mandarin Chinese [2]. Note that these tasks were performed under the spirit of ToBI [3] so that they could be compared with similar tools developed for other languages. The issues we have dealt with over time ranged from the segmental layers to the prosodic layers, the latter being our more recent focus. Our initial effort to characterize labeling system at the prosodic layer included designed a labeling system for breaks/pauses in continuous speech. The ToBI group has since acknowledged and adopted our break labeling system for Mandarin at the prosodic layer.

We noted that the organization of prosody, as with the organization of segmental phenomena, can be seen from different perspectives. Analyzing the manifestation appeared to be one from the most logical aspect if we approach the speech data from the phonetic point of view. However, though our previous approach focused on the phonetic aspects of continuous speech, that is, our analyses and experiments emphasized the physical as well as perceptual properties of speech sounds only, we have not overlooked the fact that speech output is in fact the ultimate derivation of many other parallel linguistic levels of information interacting and interplaying at the same time. In other words, any particular speech event can also be seen in that light. Take pauses/breaks for example, labeling them through a listening test means working with the physical data only and largely trying to characterize these phenomena from the perceptual point of view. In this particular chosen frame of work, arriving at perceptual consistency within and among transcribers was crucial since their labeling results would serve as the basis of the development of labeling tools. But if we take a step away from the physical data only, a reasonable question would be: Can we look beyond the lower linguistic level of phonetic information to other higher linguistic levels for more information that we know are significant, or even, for possible governing factors that to a large extent contribute to the occurrences of these breaks/pauses? In short, what could be the major governing factors of prosodic organization [4,5,6]?

The syntactic structures involved are definitely one of the most linguistically feasible factors to consider [7,8], especially with respect to syntactic boundaries. In this paper, we will report our attempt to investigate whether such interplay can be derived through correlation measurements between perceived breaks/pauses and syntactic boundaries. We will look into the prosodic manifestation and syntactic boundaries through a correlation analysis between manual labeling results with the syntactic parsing analyses.

EXPERIMENTS

Based on the assumptions that (1.) the position where breaks/pauses occur in running speech may correspond to most but not necessarily all major syntactic boundaries and (2.) breaks/pauses should not occur across syntactic boundaries, we carried out two independent kinds of labeling and later mapped the results. 1107 prosody oriented Mandarin utterances from our speech database were used. These utterances are continuous read speech by one female speaker in her late 20’s; they range from 7 to 140 syllables/characters in duration. Of these 1107 utterances, 804 are declarative utterances and 303 are exclamatory utterances in structure. The two independent kinds of labeling are: (1.) the perceived breaks that occurred in the speech data of these utterances and (2.) derived syntactic boundaries through text analysis software. To avoid inter-transcriber inconsistencies, only one trained transcriber’s labeling results were used. The transcriber labeled the breaks through listening tasks using a ToBI-spirited labeling system developed for Mandarin by our group [2]. The 5-increment 6-step break labeling system (B0 to B5) characterizes breaks as follows: B0 corresponds to reduced syllabic boundary, B1 normal syllabic boundary, B2 minor phrase boundary, B3 major phrase boundary, B4 breath-group boundary and B5 prosodic group boundary. The labeling results used represent perceived breaks by one human listener. Figure 1 shows an example of a labeled utterance for breaks perceived.

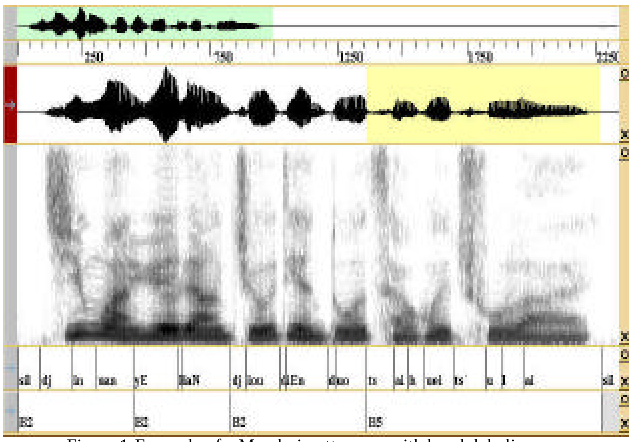

Figure 1 Example of a Mandarin utterance with break labeling

Independently, the text of the 1107 utterances was tagged for their syntactic structures using the CKIP parser [9]. The parser is essentially a lexical feature-based grammar formalism that unifies the thematic structures, phrasal structures and head-daughter dependency relations into a single uniform representation structure using a set of bracketing notations. A different transcriber for the CKIP project had to hand-tailor the parsed results to meet the needs of our purpose. Note that there are 11 possible brackets (from 0 to 10 from no bracketing up) in the tree bank structuring of the CKIP parser. 0 bracket represents no syntactic boundary, 1 bracket a phrase boundary, 2 brackets a phrase with possibly a modifier. Higher number the brackets denote more layers of bracketing, and hence where higher level of syntactic boundaries occurs. The same utterance shown in Figure 1 was parsed as:

((djin uan)|(yE lian)|(djiou diEn duo)| tsai | huei | ts`u lai)
((今晚)|(月亮)|(九　多)|才|會|出來)

We then mapped the above two independent labeling results together by tallying the number of bracketing each break label corresponds to; with the high likelihood that each break label mapped on to more than one bracket. For example, the mapping of the utterance shown above is shown as follows:

((今晚)|(月亮)|(九　多)|才|會|出來)
B2　B2　B2　B5
1　1　1　2 (# of brackets)

For each kind of break labeled, we then tallied the frequencies of its correspondence to the 11 possible brackets. The following example shows another correlation analysis of a much longer utterance without the waveform.

Parsed results for syntactic structures:

(((台電|核能|四廠)|的|發包)|已|吸引|(的((美國|(奇　|和|西屋)|公司)|的)|高度|興趣))(而|(美國|波音|公司)|也|(向|(的((我國)|的)|((中華|和|長榮)|航空|公司)))|推銷|(的(((其)|即將|問世)|的)|最新型|波音|七七七型|飛機))

Labeled results for breaks perceived:

台電 B2 核能四廠的 B2 發包 B3 已吸引 B2 的美國 B2 奇 B3 和西屋公司的 B2 高度興趣 B5 而美國 B2 波音公司 B3 也向的我國的 B2 中華 B3 和長榮 B2 航空公司 B3 推銷的其 B2 即將問世的 B2 最新型 B2 波音 B2 七七七型 B2 飛機 B5

Mapping of the above two:

(((台電|核能|四廠)|的|發包)|已|吸引|(的((美國|(奇　|和|西屋)|
B2　B2　B3　B2　B2　B3
0　0　2　0　0　0

公司)|的)|高度|興趣))(而|(美國|波音|公司)|也|(向|(的((我國)|
B2　B5　B3　B3
3　5　0　1

的)|((中華|和|長榮)|航空|公司)))|推銷|(的(((其)|即將|問世)|
B2　B3　B2　B3　B2
2　0　1　4　1

的)|最新型|波音|七七七型|飛機))
B2　B2　B2　B2　B5
3　0　0　0　5 (# of brackets)

RESULTS

Tables 1a and 1b summarizes the results of correlation by utterance types.

Declarative				
breaks / # of brackets	B2	B3	B4	B5
0	35.8%	10.7%	5.5%	0.4%
1	31.0%	16.6%	3.2%	2.7%
2	19.6%	29.3%	24.7%	28.0%
3	8.9%	21.6%	23.6%	22.1%
4	3.3%	13.9%	20.7%	22.5%
5	0.9%	5.6%	11.5%	14.4%
6	0.2%	1.8%	7.2%	5.2%
7	0.1%	0.4%	3.2%	3.0%
8	0.0%	0.2%	0.0%	0.8%
9	0.0%	0.0%	0.6%	0.7%
10	0.0%	0.0%	0.0%	0.1%
# of tagged tokens	5341	1885	348	825

Table 1a. Correlation between perceived breaks and corresponding syntactic boundaries in declarative utterance

Exclamatory				
breaks / # of brackets	B2	B3	B4	B5
0	34.8%	4.7%	1.3%	0.3%
1	36.3%	15.9%	8.9%	15.5%
2	18.0%	36.0%	36.7%	25.7%
3	7.6%	22.7%	29.1%	26.0%
4	2.8%	12.5%	8.9%	16.9%
5	0.3%	6.4%	7.6%	9.8%
6	0.1%	1.5%	7.6%	4.1%
7	0.0%	0.2%	0.0%	1.4%
8	0.0%	0.0%	0.0%	0.3%
9	0.0%	0.0%	0.0%	0.0%
10	0.0%	0.0%	0.0%	0.0%
# of tagged tokens	1369	528	79	296

Table 1b. Correlation between perceived breaks and

corresponding syntactic boundaries in exclamatory utterance

Table 1a shows that in the 804 declarative utterances, most of B2 (minor phrase boundary) correspond to two kinds of brackets, namely, 0 bracket (35.8%) and 1 bracket (31%). The latter would be the most reasonable correspondence, the former may attribute to speaker's intension and/or some kind of emphasis. Note that the next highest correspondence occurred for 2 brackets (19.6%). B3 (major phrase boundary) corresponded mostly to 2 brackets (29.3%) and 3 brackets (21.6%). Note also that 13.9% of B3 corresponded to 4 brackets. B4 (breath-group boundary) corresponded mostly to 2 brackets (24.7%), 3 brackets (23.6%) and 4 brackets (20.7%) whereas B5 (prosodic group boundary) corresponded mostly to 2 brackets (28%), 3 brackets (22.1%) and 4 brackets (22.5%). Note that 11.5% of B4 corresponded to 5 brackets whereas 14.4% of B5 also corresponded to 5 brackets.

Table 1b shows that in the 303 exclamatory utterances, most of B2 (minor phrase boundary) correspond to two kinds of brackets, namely, 0 bracket (34.8%) and 1 bracket (36.3%). Note that the next highest correspondence occurred for 2 brackets (18%). We also note that B2 behaved somewhat similarly for both declarative and exclamatory utterances. B3 (major phrase boundary) corresponded mostly to 2 brackets (36%) and 3 brackets (22.7%). Note that 15.9% of B3 corresponded to 1 bracket while 12.5% of B3 corresponded to 4 brackets. In other words, B3 is seen to correspond differently to numbers of syntactic bracketing in declarative and exclamatory utterances. B4 (breath-group boundary) corresponded mostly to 2 brackets (36.7%) and 3 brackets (29.1%) and 4 brackets (20.7%) whereas B5 (prosodic group boundary) corresponded mostly to 2 brackets (25.7%) and 3 brackets (22.1%) and 4 brackets (26.1%). Note that 15.5% of B5 corresponded to 1 bracket and 16.9% of B5 corresponded to 4 brackets.

DISCUSSION

Though we were able to obtain some kind of correlations, they are by no means high. However, we believe that we should be able to improve the correlation later on. As mentioned above, the CKIP parser is not only a syntactic parser, it also included lexical features and thematic structures. The latter may contribute to the high number of brackets designed. Secondly, 11 layers of bracketing may be too many for our kind of correlation, therefore breaking up the correspondence. One direction would be to collapse the number of brackets into larger syntactic categories in our next step, thereby reducing the steps from 11 to less numbers. The rationale would be that we would be interested in looking into larger syntactic units only, such as noun phrases (NP), verb phrases (VP) and so forth in the purely syntactic sense. In other words, given the amount of manually labeled prosodic data available, we should probably aim at fine-tuning the syntactically parsed results for possible better correlation. Thirdly, the unexpected occurrence of B5 (prosodic group boundary), the longest break/pause, in the middle of a long utterance is worth more detailed investigation. We reason here that breaks could be used together with prominence, another layer of prosodic property currently under investigation. We intend to integrate systematically labeled prominence information in our next project, too.

CONCLUSION

In this study, we reported our first attempt to correlate the relationship between syntactic structure and actual prosodic manifestation in continuous speech of Mandarin Chinese, using perceived breaks/pause as our prosodic reference. Independent labeling/tagging results of the two linguistic levels were mapped to test if correlation could be found. We also see the reported attempt as a first step towards a working model for the organization of prosody. The linguistic significance here is quite clear: the interaction between breaks/pauses and syntactic boundaries can be seen through this kind of investigation. A speech database certainly offered a much wider range of evidence that furthered our knowledge in this respect. Our preliminary results showed that mapping was indeed found. However, if the proposed type of correlation were to be utilized for application in speech science research or tools, finer tailoring of the mapping would be necessary. Nevertheless, direct application to speech synthesis would be a highly likely area, where speech recognition could also benefit from the kind of results obtained. We believe that the incorporation of more levels of prosodic information such as prominence and emphasis would be necessary and crucial to develop better tools. Our future work is geared towards that direction.

REFERNCES

[1] Tseng, C. "Investigating Mandarin Chinese prosody through speech database". *Proceedings of the 1999 Oriental COCOSDA Workshop*, Taipei, Taiwan, pp. 141-144, 1999.

[2] Tseng, C. & Chou, F. "A Prosodic labeling system for Mandarin Speech Database". *Proceedings of the XIVth International Congress of Phonetic Sciences*, San Francisco, CA., pp. 2379-2382, 1999.

[3] Pitrelli, J., Beckman, M. & Hirschberg, J. "Evaluation of prosodic transcription labeling reliability in the ToBI framework". *Proceedings of the third international conference on spoken language processing*, Yokohama, Japan, 2, 123-126, 1994.

[4] Hirshchberg, J. & Pierrehumbert, J. " The intonational structuring of discourse". *Proceedings of the 24th annual meeting of the Association for Computational Linguistics*, New York, NY pp. 136-144, 1986.

[5] Thorsen, N. "Intonation contours and stress group patterns in declarative sentences of varying length in ASC Danish". *Annual report of the Institute of Phonetics*, University of Copenhagen, 14, pp. 1-29, 1980.

[6] Thorsen, N. "Intonation contours and stress group patterns in declarative sentences of varying length in ADC Danish — supplementary data". *Annual report of the Institute of Phonetics*, University of Copenhagen, 15, pp. 13-47, 1981.

[7] Lehiste, I. "The timing of utterances and linguistic boundaries". *Journal of the Acoustical Society of America*, 51, pp. 2018-2024, 1972.

[8] Mertens, P. "Intonational grouping, boundaries and syntactic structure in French". In *ESCA Workshop on Prosody, Lund Working Paper*, 41, pp. 155-159, 1993.

[9] Chen, K., Luo, C. Gao, Z., Chang, M., Chen, F., Chen, C. & Huang, C., "The CKIP Chinese treebank: Guidelines for annotation". http://godel.iis.sinica.edu.tw/CKIP/treebank/, 2000.