What Do Speakers Do and Why—
The Story of Prosody-Syntax Non-Overlap and Higher Level Discourse Information

Chiu-yu Tseng and Zhao-yu Su

Phonetics Lab, Institute of Linguistics, Academia Sinica, Taipei, Taiwan
cytling@sinica.edu.tw

Abstract

We have previously addressed the functions of Mandarin fluent speech prosody from a top-down perspective in light of higher level discourse information and cross-phrase prosodic associations. Postulating a prosody hierarchy HPG (Hierarchical Phrase Grouping) of multi-phrase speech paragraphs by systematic treatments of boundaries and breaks as discourse related, we were able to quantitatively account for higher level contributions in the cross-phrase prosody context by acoustic parameters, and explain how such association triggers lower level nodes to modify systematically (Tseng et al, 2004; 2005; 2006). In this paper, we further investigate within-paragraph prosody-syntax non-overlaps to look for contributions of higher information rather than on how prosody disambiguates underlying syntactic structures most noted in the literature [1, 2, 3, 4, 5]. We hypothesize that most of such non-overlaps are due to higher-level information and CNA be accounted for. We define overlap by mapping of annotated boundaries in speech corpora and punctuation marks (PM) in corresponding text; whereas non-overlaps are where (1.) no PM in text but a boundary is tagged in speech data, (2.) a PM in text but no boundary occurs in speech data and (3.) mismatch between PM and produced boundary. Three types of speech corpora differing in style and format were used: (1.) reading of 26 discourse pieces up to 900 more syllables/characters by 2 radio announcers of plain text (CNA), (2.) reading of weather forecast by 2 untrained speakers (WF) and (3.) reading of three Chinese Classics in three different rhyming formats by 2 untrained speakers (CL). We noted that speakers do respect PMs as indication of syntactic structure since only a small portion of PMs (commas in particular) were overlooked in speech data (4.3%, 4.41% in CNA; 13.41%, 8.50 in WF; 6.50, 2.17% in CL). However, we note considerable higher percentage of inserted PPh boundaries (and pauses) exist in speech data where no corresponding PMs occur in text (26.56%, 30.49% in CNA; 23.08%, 20.19% in WF; and 11.69%, 3.14% in CL). We also find relatively high between-speaker overlaps exist in these non-overlaps (45.15%, 37.02% in CNA; 36.57%, 39.20% in WF; and 8.33%, 30.00% in CL) indicating such non-overlaps are by no means random. These non-overlaps are analyzed by syntactic structure and by paragraph positions using quantitative methods to demonstrate contributions from higher level paragraph information.

Index Terms: Hierarchical Prosody Group, HPG, discourse prosody, higher level contribution, prosody-syntax non-overlap

1. Introduction

We have established previously [6, 7, 8] from quantitative corpus analyses of Mandarin Chinese that fluent speech prosody contains higher level discourse information above intonation unit (IU). IU functions as sub-prosody unit in spoken discourse and is subject to change by paragraph specifications. We further stated that higher information is the semantics that associates phrases and sentences into coherent speech paragraphs beyond syntax government, delivered through cross-phrase prosodic context, most notably as intonation variations. Our Hierarchical Prosodic phrase Grouping (HPG, formerly termed PG) framework specifies how by three paragraph positions –initial, -medial and –final, higher level discourse information constrains and triggers individual phrase intonations to adapt systematically in order to yield multi-phrase paragraph prosody. The super-positioning of layered prosodic information of various domain collectively contributes to output prosody; the make-up of layered contributions could be accounted for quantitatively. As a result, we argue that paragraph flow could be represented by cross-phrase prosody deep structure; output intonation variations are systematic and predictable. (See details in [7, 8].) The three relative HPG-positions -initial, -medial and –final define phrase units into paragraph roles in paragraph and cross-phrase dynamics are simply how IU must bear the beginning, on-going and terminating functions. Thus, systematic multi-phrase templates of F0 contour patterns, syllable duration adjustments, intensity distributions and boundary break patterns could be quantitatively derived from corpus analyses of speech data. Correlating modular acoustic simulation models were also constructed [8]. Figure 1 shows the 6-layer tree diagram of the HPG framework in prosodic units that accounts for multi-phrase output prosody. From bottom up, the layered nodes are syllables (SYL), prosodic words (PW), prosodic phrase (PPh), breath groups (BG), prosodic phrase groups (PG) and Discourse. The upper prosodic layers/levels above PPh CNA also collapse to accommodate discourse of various lengths.

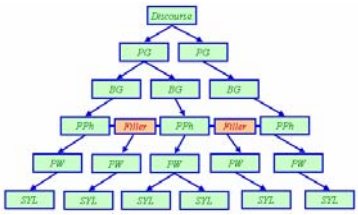


Figure 1. A schematic tree diagram of phrase-grouping discourse organization in prosodic levels and units, including between-phrase fillers and markers

Our current hypothesis is that in addition to speaker intension, within-paragraph prosody-syntax non-overlaps are mostly higher information related. More understanding of what speakers actually do when producing fluent speech reveals information fluent speech prosody bears. Since paragraph and discourse involve semantic cohesion above sentences, more information than syntactic governing exists in speech flow and prosody functions much more than disambiguating underlying syntactic structures [1, 2, 3, 4, 5]. We believe one feasible way to look into this aspect is to perform syntactic analyses of text data for nodes and boundaries [9, 10] and subsequently compare with actual speech data [11] to look into the speech-syntax non-overlap. In other words, what speakers actually do and why they did it.

We hypothesize that most of such non-overlaps are due to higher-level information and CNA be accounted for. We define overlap by mapping of annotated boundaries in speech corpora and punctuation marks (PM) in corresponding text; whereas non-overlaps are where (1.) no PM in text but a boundary is tagged in speech data, (2.) a PM in text but no boundary occurs in speech data and (3.) mismatch between PM and produced boundary. Three types of speech corpora differing in style and format were used: (1.) reading of 26 discourse pieces up to 900 more syllables/characters by 2 radio announcers of plain text (CNA), (2.) reading of weather forecast by 2 untrained speakers (WF) and (3.) reading of three Chinese Classics in three different rhyming formats by 2 untrained speakers (CL).

2. Text and Speech Data

Three types of corpus and corresponding Mandarin speech data are used to examine prosody-syntax non-overlap. The three types of corpus contain (1.) reading of plain text of 26 discourse pieces by 2 radio announcers (one male and one female) (CNA), (2.) reading of weather forecast by 2 untrained native speakers (WF) and (3.) reading of three formats of Chinese Classics by 2 untrained speakers (CL). The location and type of punctuation marks in text reveal syntactic structures of corpus. What is of interest to us whether the punctuation marks involve semantic cohesion above sentences that signify larger semantic units. The distribution of each type of punctuation marks by corpus is listed in Table 1.

Table1. Distribution of punctuation marks in text by corpus type.

Corpus / # of PM	1.CNA	2.WF	3.CL
Comma ，	744	427	269
period。	317	110	190
Pause 、	92	83	13
Semicolon ；	20	18	8
Exclamation ！	14	2	14
Question ？	44	0	10

All corresponding speech data are reading of the above text produced in sound proof chambers. Pre-analysis annotation included automatically labeled segmental identities by the HTK toolkit in SAMPA-T notation, followed by subsequent manual tagging of perceived boundary breaks by trained transcribers using the Sinica COSPRO Toolkit [12]. Annotation results were spot-checked by professional transcribers for segmental alignments. Table2 summarizes speech data corresponding to three types of text corpus.

Table2. Summary of speech data by corpus type

Discourse	speaker	# of Syl	# of PPh	# of Discourse	speech rate (ms)/Syl
CNA	f051	11592	1092	26	200
	m051	11600	1207	26	189
WF	f054	7054	676	34	193
	m054	7096	728	34	165
CL	f054	3502	308	26	271
	m056	3510	318	26	202

3. Method of Analysis

The purpose of the analysis is to examine the proportion of non-overlaps and look for patterns that CNA be used to model the probability of such non-overlaps. All positions of PMs are correlated with perceptual labels to find out the corresponding perceptual boundaries in speech corpora. Figure 2 shows the PM-boundary correlation.

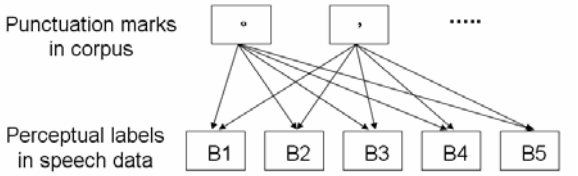


Figure 2: Correlation of annotated boundaries in speech corpora and punctuation marks (PM) in text.

Initial investigations focus on PM analyses in relation to prosody boundaries. Subsequent examinations will aim at within-phrase syntactic analyses.

4. Results

Results in percentage are presented to compare the distribution patterns among different types of corpus and among different speakers.

4.1. Distribution of annotated boundary breaks in relation to each type of PM

Preliminary syntactic analyses of text data by PM are performed to examine the matching distribution of annotated boundaries in actual speech data. Results are summarized in Table 3. The blue blocks in Table 3 denote syntax-prosody overlaps and their respective distributions, i.e., PM in text vs. matched boundary breaks in speech data. On the other hand, prosody-syntax non-overlaps are defined as mismatches; their respective distributions presented in red.

Table3. Distribution of annotated boundary breaks in relation to each type of PM used in Chinese text. "、" denotes a slight-pause mark used to set parallel words or short phrases;

" ! " exclamation mark, " ; " semicolon, " , " comma, and " 。" period.

CNA

f054	B1	B2	B3	B4	B5
、	7.23%	60.24%	32.53%	0.00%	0.00%
!	0.00%	0.00%	100.00%	0.00%	0.00%
;	0.00%	0.00%	33.33%	55.56%	11.11%
,	0.00%	5.62%	72.13%	21.08%	1.17%
。	0.91%	0.00%	16.36%	30.91%	51.82%

m054	B1	B2	B3	B4	B5
、	7.23%	20.48%	72.29%	0.00%	0.00%
!	0.00%	0.00%	0.00%	50.00%	50.00%
;	0.00%	0.00%	66.67%	33.33%	0.00%
,	0.00%	5.15%	81.03%	12.88%	0.94%
。	0.00%	0.00%	11.82%	20.00%	68.18%

WF

f051	B1	B2	B3	B4	B5
、	2.17%	4.35%	91.30%	2.17%	0.00%
!	0.00%	0.00%	50.00%	14.29%	35.71%
;	0.00%	0.00%	55.00%	20.00%	25.00%
,	0.13%	3.36%	74.87%	20.16%	1.48%
?	0.00%	0.00%	34.09%	20.45%	45.45%
。	0.00%	0.00%	25.55%	38.17%	36.28%

m051	B1	B2	B3	B4	B5
、	3.33%	11.11%	83.33%	2.22%	0.00%
!	0.00%	0.00%	50.00%	28.57%	21.43%
;	0.00%	0.00%	70.59%	23.53%	5.88%
,	0.92%	4.19%	76.31%	17.41%	1.18%
?	0.00%	2.33%	46.51%	39.53%	11.63%
。	0.32%	0.00%	33.44%	32.18%	34.07%

CL

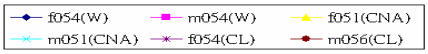
f054	B1	B2	B3	B4	B5
、	0.00%	38.46%	61.54%	0.00%	0.00%
!	0.00%	0.00%	21.43%	35.71%	42.86%
;	0.00%	0.00%	37.50%	62.50%	0.00%
,	0.00%	10.41%	85.13%	4.46%	0.00%
?	0.00%	0.00%	50.00%	50.00%	0.00%
。	0.00%	0.00%	11.05%	60.00%	28.95%

m056	B1	B2	B3	B4	B5
、	15.38%	15.38%	69.23%	0.00%	0.00%
!	0.00%	0.00%	14.29%	42.86%	42.86%
;	0.00%	0.00%	50.00%	50.00%	0.00%
,	0.00%	2.60%	93.31%	4.09%	0.00%
?	0.00%	0.00%	50.00%	50.00%	0.00%
。	0.00%	0.00%	17.37%	61.05%	21.58%

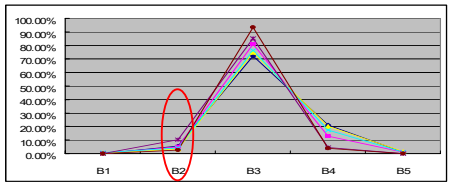
4.2. Comparison of distribution of annotated boundaries by PM

The distributions of annotated boundaries in speech data with respect to PM in text are also compared. Results are presented in Figure 3.

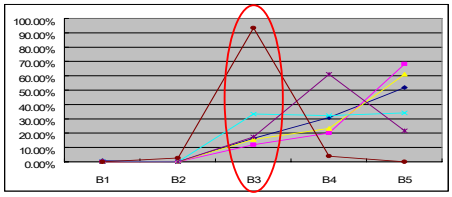
The behavior of each speaker is plotted in different colors; each trajectory in Figure 3 thus denotes distribution patterns by speaker. PMs with too few samples as shown in Table1 were not included for comparison. Prosody-syntax non-overlaps are presented in red ellipses.



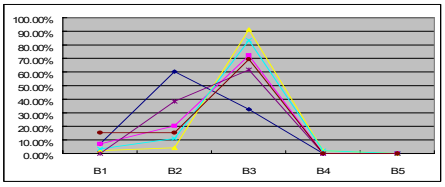
Comma- ,



Period - 。



Slight-pause between parallels - 、



Semicolon - ;

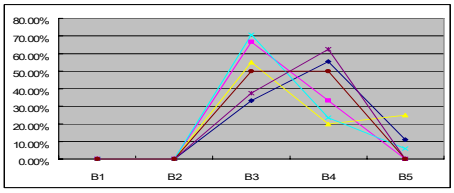


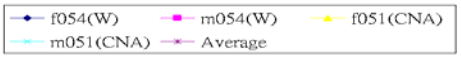
Figure 3: Distributions of annotated boundary breaks with respect to PM in text.

4.3. Classification of speech data

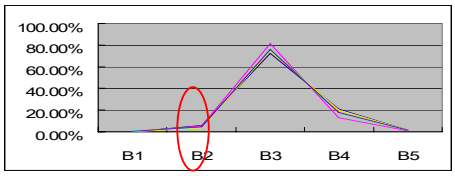
In a previous prosody study of reading Chinese rhymed classics differing in degrees of regularity of employed rhyming template, we found that speakers' production planning is conditioned by the degrees of template regularities [13]. The more regular the rhyming templates are; the larger the planning units become. In other words, speakers' planning strategies fine-tuned systematically by the nature of materials to be read, and the prosody outputs vary. Thus we categorize the speech data used in the present data as a control. The speech data were classified as colloquial speech with no built-in rhyming templates (CNA&WF) and reading of Chinese classics with varied rhyming templates (CL) to separate colloquial speech from more planned reading. Our aim is to find general probability model for the two types of speech data by different prosody format.

4.3.1. Distribution of boundary breaks by PM and prosody format CNA&WF

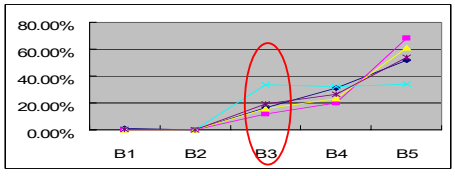
In the colloquial speech data (CNA&WF), all four speakers exhibited similar patterns of distributions in terms of boundaries and pauses by PMs, comma or period. Three out of the four speakers also exhibited similar patterns for pause mark and semicolon, as the similar trajectories shown in Figure 4. Thus the similar distribution patterns in the majority of speakers CNA be regarded as the general probability model for specific PM. However, speaker difference does exist as observed in speaker f054 (W) who has exhibited semicolon- and pause-mark- patterns different pattern from the rest of the speakers.



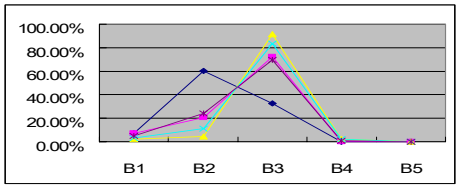
Comma- ，



Period - 。



Slight-pause between parallels - 、



Semicolon- ；

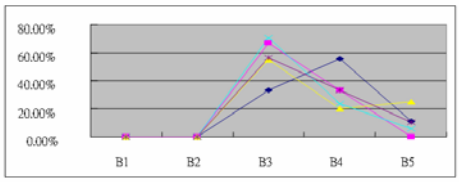
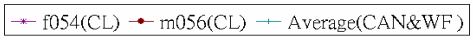


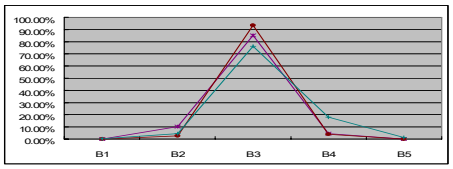
Figure 4: Distribution of annotated boundary breaks corresponding to specific PM in CNA&WF

4.3.2. Comparison of boundary break distribution by PM and prosody formats

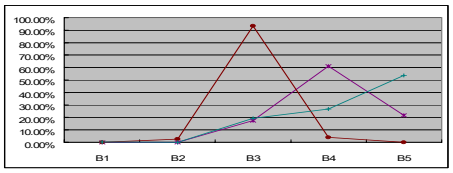
The distribution of boundary breaks in relation to PM for both colloquial speech CNA&WF and Chinese classics CL is presented in Figure 5. The average distribution of four speakers in CNA&WF is calculated and used to compare the distributions of two speakers in CL. Results show that except for PM comma, distribution patterns in CL are quite different from the pattern showed in CNA&WF.



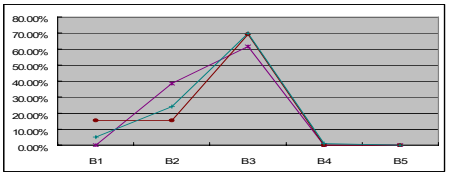
Comma - ，



Period- 。



Slight-pause between parallels - 、



Semicolon- ；

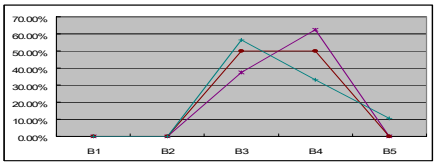


Figure 5: Distribution of annotated boundary breaks corresponding to specific PM in CL.

5. Discussion

We hypothesized that due to discourse information which constrains higher-level cohesion, syntax-prosody non-overlaps are bound to occur in fluent speech and shown through prosody. However, such non-overlaps are recoverable and thus can be accounted for. We define overlap by mapping of

annotated boundaries in speech corpora and punctuation marks PM in corresponding text; whereas non-overlaps are where (1.) no PM in text but a boundary is tagged in speech data, (2.) a PM in text but no boundary occurs in speech data and (3.) mismatch between PM and produced boundary. Three types of speech corpora differing in style and format were used. Distribution of annotated boundary breaks in relation to each type of PM showed a high percentage of speaker overlap (see Figure 4). Comparison of distribution of annotated boundaries by PM also indicates cross-speaker patterns of non-overlaps (see Figure 4) are rather consistent. Dividing the speech data by prosody format revealed different planning strategies across speakers. When reading plain text unprepared in colloquial speech, speakers respect phrase and sentence boundaries indicated by PMs, comma and period (Figure 4), but freely interpret PMs that indicate smaller and/or parallel units as indicated by pause mark and semicolon. In other words, the non-overlaps exist in smaller units. Interesting enough, when reading text of varied degrees of rhyming regularity, the reverse pattern emerged (Figure 5). The speakers are well aware and to some extent knowledgeable of the templates used; the domain of cross-phrase look-ahead increased. As a result, PM period indicating sentence boundary is sometimes overlooked, causing the non-overlaps to occur. Sentences are treated more like phrases, discourse associations prevail, and the entire discourse becomes a complex sentence, especially when it is relatively short.

We further classified speech data CNA and WF as colloquial speech with no built-in rhyming templates and CL as rhymed speech of varied rhyming templates to bring out the prosody differences. As shown in Table4, we noted that speakers do respect PMs as indication of syntactic structure since only a small portion of PMs (commas in particular) were overlooked in speech data (4.3%, 4.41% in CNA; 13.41%, 8.50 in WF; 6.50, 2.17% in CL).

Table4. Distribution of PMs overlooked in speech data

Corpus	speaker	overlooked PMs	overlooked ratio
CNA	f051	56	4.30%
	m051	57	4.41%
WF	f054	87	13.41%
	m054	54	8.50%
CL	f054	33	6.50%
	m056	11	2.17%

However, we note considerable higher percentage of inserted between-phrase boundaries B3 (and pauses) exist in speech data where no corresponding PMs occur in text as shown in Table5 (26.56%, 30.49% in CNA; 23.08%, 20.19% in WF; and 11.69%, 3.14% in CL).

Table5. Distribution of B3s in speech data where no corresponding PMs occur in text

Corpus	speaker	# w/out corresponding PMs	ratio w/out corresponding PMs
CNA	f051	290	26.56%
	m051	368	30.49%
WF	f054	156	23.08%
	m054	147	20.19%
CL	f054	36	11.69%
	m056	10	3.14%

Preliminary syntactic analyses of these within-phrase boundaries revealed boundaries of syntactic nodes as expected [1], and can be used as evidence of the speaker's on-line parsing at the syntactic level. Relatively high between-speaker overlaps are found in these non-overlaps (45.15%, 37.02% in CNA; 36.57%, 39.20% in WF; and 8.33%, 30.00% in CL) as shown in Table6, indicating such non-overlaps are by no means random.

Table6. Distribution of between-speaker overlaps of inserted between-phrase boundaries B3

corpus	speaker	# of between-speaker overlaps	ratio of between-speaker overlaps
CNA	f051	107	45.15%
	m051		37.02%
WF	f054	49	36.57%
	m054		39.20%
CL	f054	3	8.33%
	f056		30.00%

6. Conclusion

We have shown from the above initial corpus investigation of syntax-prosody non-overlaps are due to on-line parallel syntactic processing at the sentence level and paragraph segmentation at the discourse level. What can not be accounted for by syntactic analyses could relatively easily be traced to higher-level discourse information instead of random variation. We believe the syntax-discourse interaction finds support from the present study; higher level paragraph discourse association must be taken into account in fluent speech prosody, and discourse information merits due attention in prosody investigations. With further and future quantitative analyses of distribution accounts from interacting information involved, we believe both natural language processing and technology development could benefit from more information at the discourse level.

7. References

[1] Lehiste, Ilse. "Phonetic Disambiguation of Syntactic Ambiguity", *Journal of the Acoustical Society of America*, Vol. 53: 1: 380, January 1973.

[2] Cooper, William E and Paccia-Cooper, Jeanne. *Syntax and Speech*, Harvard University Press, 1980.

[3] Price, P. J. Ostendorf, M. Shattuck-Hufnagel, S. and Fong, C. "The use of prosody in syntactic disambiguation**",** *Journal of the Acoustical Society of America*, Vol. 90:6:2956-2970, December 1991.

[4] Beach, C. M. "The interpretation of prosodic patterns at points of syntactic structure ambiguity: evidence for cue trading relations", *Journal of memory and language*, vol. 30:6:644-663, 1991.

[5] Nakatani, C. and Hirschberg, J. "A speech-first model for repair detection and correction", *Proceedings of the 31st annual meeting on Association for Computational Linguistics* , 46 -53, Columbus, Ohio, 1993.

[6] Tseng, C. "Prosody Analysis", *Advances in Chinese Spoken Language Processing*, World Scientific Publishing, Singapore: 57-76, 2006.

[7] Tseng, C. Pin, S. and Lee, Y., Wang, H. and Chen, Y. "Fluent Speech Prosody: Framework and Modeling", *Speech Communication, Special Issue on Quantitative Prosody Modeling for Natural Speech Description and Generation*, Vol. 46:3-4: 284-309, 2005.

[8] Tseng, C and Lee Y. “Speech rate and prosody units: Evidence of interaction from Mandarin Chinese”, *Proceedings of the International Conference on Speech Prosody 2004*, 251-254, 2004.

[9] Jurafsky, D. and H.Martin, J. *Speech and Language Processing*, Prentice Hall Publishing U.S.A, 2000.

[10] Chen, K. Tseng, C and Tai, C. "Predicting Prosody from Text", *5th International Symposium on Chinese Spoken Language (ISCSLP 2006).* Kent Ridge, Singapore, 2006.

[11] Tseng, C. and Chen, D. "The Interplay and Interaction Between Prosody and Syntax: Evidence from Mandarin Chinese", *6th International Conference on Spoken Language Processing (ICSLP 2000).* Vol.II , 95-97. Beijing, China, 2000.

[12] Sinica COSPRO and Toolkit: http://reg.myet.com/registration/corpus/en/Main.asp

[13] Tseng, C and Su, Z. "From One Base Form to Multiple Output Styles-- Predicting Stylistic Dynamics of Discourse Prosody", *Interspeech 2007 Eurospeech.* 110-113. Antwerp, Belgium, 2007.