

Prosodic prompts and information planning units in continuous speech— Relative allocation and compensation of prosodic highlight

Helen Kai-yun Chen, Wei-te Fang, Chiu-yu Tseng

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

Abstract

Based on the assumption that prosodic highlight allocation in continuous speech directly reflects speakers' information planning, this study explores perception-based prominences for indexing *key* information [KEY], as well as cuing the up-coming key information, namely *projector* [PJR] with its trajectory of *projection* [PJN], functioning to indicate information planning regardless of the new-vs-old status. By defining both [KEY] and [PJR]-[PJN] as basic units of information planning, the study focuses on the allocation of and compensations between both units in continuous speech of diverse genres, and makes clear the significance of advanced prosodic prompting in relation to information planning. On the one hand, by calculating emphasis density throughout the trajectory of both units, a consistent pattern of heavy-to-light information loading from the beginning of the trajectory is identified. On the other, through distributions at higher-level discourse prosodic units, we test the hypothesis of the correlation between the location of both units and their trajectory size. Specifically, the results confirm that the larger the [PJR]-[PJN] size, the earlier positioned the beginning and later the ending of the unit would be at the breath-group (BG) level. Eventually the findings contribute to a comprehensive account for the mechanism behind information planning in continuous speech via the relative allocation and compensation between the two major information planning units.

Keywords: prosodic highlight prompting; projector-projection as information planning units; key information and projection; continuous speech information planning

1. Introduction

In continuous speech, pitch accents play the major role of signaling the information reflected from the unit which speakers assign accents to (cf. [1]). From the viewpoint of perception, the more prominently perceived parts of the speech are usually associated with the marking of either the focal or new information, as discussed in ([2]; [3]) from prosodic phonology and also in speech prosody. However, what has been less addressed till very recently is the function of prosodic highlight for indexing 'specific parts of the discourse' (i.e. [1]) in order to generate anticipated prediction and expectation of the *projection* for the up-coming information in the planning of the on-going speech ([4]; [5]). In other words, prominence placement within the speech flow is *not* limited to the marking of salient information. Instead, it is held that one crucial aspect of the continuous speech planning involves how speakers deploy and compensate the prosodic highlight for indexing both the focal information and the anticipated projection of the upcoming information. Together, both indexing in terms of prosodic highlight arrangement are the key to what makes the communication effective.

We have identified recently (i.e. [4]) that almost 70% of the annotated perceived emphasis tokens within continuous speech are either indexes of *key information* [KEY] or *projector* [PJR] of the up-coming information. It has also been noted that the use of [PJR] outnumbers that of [KEY] by about 15-25% ([4]). Such results suggest that the functions of [PJR] as an information unit deem more attention. In order to provide a comprehensive account toward the scheme of on-line information planning in continuous speech, we therefore compare the above 2 major types of information units across 4 diverse genres of continuous speech. As will be shown, information planning is indeed directly reflected from the allocation and compensation between the prosodic highlight in association with not only [KEY] as the focal information, but also [PJR] as projecting in advance the soon-to-arrive key information.

1.1. Information planning units: [KEY] and [PJR]-[PJN]

For the current study, [KEY] as a perceived prosodic highlight prompted unit of information planning is identified, based on the correspondence of the focal, most salient, and at times new information. [PJR] as another major unit of information planning, on the other hand, is prompted also by perceived prosodic highlight but with an alternative function to create in advance expectations for the up-coming of the key information. In this latter case, it is crucial to associate the prosodic highlight prompted [PJR] with its respective range of intended information planning, i.e. the *projection* [PJN], which includes the soon-to-arrive key information. Here *projection* is defined as a perceptible completion in anticipation which can be syntax- and/or semantics-based, and which also coincides sometimes with a prosodic completion (cf. [6]). In fact, in related literature the idea of projection can be traced back to the discussion of projection principle from theoretical syntax ([7]; [8]). Moreover, the exploration of projection has also been found in studies from the interaction perspective ([9]; [10]). We note, however, that syntactically defined projection is applicable to syntactic phenomena that only account for phrase/sentence-level of planning, and often does not translate well to the planning of the continuous speech ([5]). As for the projection within the interaction-based studies, although there has been discussions based on the interaction perspective and syntax or even prosody, most attention, however, has been placed on only features at the projection completion (cf. [6]). As explicated in [6], these features may occur too late to account for the anticipated planning of the projection. After all, listeners could not and would not wait until the projection completion to begin with the on-line information processing ([5]).

When incorporating the idea of projection in the current study, we start from the prosodic highlight prompted [PJR] and pay attention to the allocation of prosodic highlight tokens throughout the perceived *trajectory* (cf. [5]) of its

respective [PJN]. In defining the relationship between the two types of information planning units, it is held that [KEY] itself pertains to a complete projection; whereas *projector* followed immediately by its respective projection forms a unit of [PJR]-[PJN] and covers at least a piece of focal information in the projection trajectory (evidences to the claim of treating [PJR]-[PJN] as a unit will be provided in Section 4.1). Specifically, here we examine the allocation of prosodic highlight within both [KEY] and [PJR]-[PJN] units at different levels of discourse-prosodic units in reflecting the deployment as well as compensation of information planning in continuous speech.

1.2. Research objectives and hypothesis

The objective of this study, therefore, focuses on the allocation of perceived prosodic highlight within the trajectory of both the [KEY] and [PJR]-[PJN] units in corresponding to the information planning in continuous speech. As explained, we examine both types of information units by various levels of discourse-prosodic units (DPU) from a hierarchical framework. At the lower *prosodic phrase* level PPh, we calculate the **emphasis density** across the trajectory of both units. At the higher discourse-based levels, we test the hypothesis if the location of the units at these levels might be correlated with the planning size of the units. As will be shown, the calculation of emphasis density demonstrates that a consistent pattern of **heavy-to-light** density across both types of units can be identified. Moreover, the correlation between the location of the [PJR]-[PJN] unit and its size in terms of information planning is further confirmed, at least at the **BG** level; while the [KEY] unit distributes evenly across the higher discourse-prosodic levels. Eventually results from the cross speech genre comparison enlighten our understanding toward the deployment and compensation of information units in the planning of the continuous speech and hence foregrounding the significance of prosodic highlight for information prompting.

2. Speech data and annotations

2.1. Speech data

For following analyses, data of 4 diverse speech genres, i.e. two of read speech and two spontaneous speech, all in Mandarin, are incorporated in this study. The read speech includes data selected from tasks of prose reading (CNA) and weather broadcast simulation (WB). As for spontaneous speech, one is a university classroom lecture (SpnL) and the other a spontaneous interaction of casual style (SpnC). Table 1 summarizes the total amount of speech data from each genre.

Table 1: Summary of total time and number of syllable of the data from 4 speech genres.

Corpora	Total time (min)	Total number of Syl
CNA	50	22988
WB	28	14083
SpnL	145	33306
SpnC	54	10756

2.2. Data preprocessing and annotations

For preprocessing, the aforementioned data have first undergone force alignments by the HTK Toolkit, and then the output was manually checked by trained transcribers. The next step involves labor-intensive annotations across all the selected data in separate layers for the following information.

2.2.1. Annotations for discourse-prosodic unit (DPU)

First of all, the speech data have been annotated for levels of discourse-prosodic units (DPU), following the hierarchical prosodic phrase grouping (HPG) proposed by [11], [12], and [13]. The HPG framework includes 5 DPU levels, marked from B1 through B5 that correspond respectively to: *syllable* (SYL), *prosodic word* (PW), *prosodic phrase* (PPh), *breath group* (BG, a physio-linguistic unit constrained by change of breathe while speaking continuously) and *multiple phrase speech paragraph* (PG) [11]. By default the boundary breaks, prosodic units and their relationship within the HPG framework can be accounted for by the following: SYL/B1<PW/B2<PPh/B3<BG/B4<PG/B5 [13].

2.2.2. Annotations for perceived prosodic highlight

The same speech data were manually annotated, in a separated layer, into a string of perception-based emphasis/non-emphasis tokens (ETs). The annotation for prosodic highlight is based on 4 relative degrees of perceived strength by prominences, following the definitions:

- E0 -- reduced pitch, lowered volume, and/or contracted segments
- E1 -- normal pitch, normal volume and clearly produced segments
- E2 -- raised pitch, louder volume and irrespective of the speaker's tone of voice
- E3 -- higher raised pitch, louder volume and with the speaker's change of tone of voice

By this annotation scheme, we emphasize the fact that the distinctions in prominences can be perceived consistently by only limited numbers of contrastive levels. It is further noted that among the 4 speech genres, only spontaneous speech data were annotated for reduction E0. The read speech data were not tagged for E0, as it is assumed that speakers rarely carry out reduction in the reading tasks.

2.2.3. Annotations for prosodic prompted information content

In a 3rd layer, we categorize the ETs with actual emphases, namely those of E2 and E3, based on the corresponding information content of each token by PW (B2). Two major categories for information content, as explained, are the *key information* [KEY] and *projector* [PJR]. After the categorization, we further annotate to each [PJR] its respective *projection* [PJN] by the perceived trajectory range in yet a separate layer. It is further noted that the prosodic prompted [PJR] is followed immediately by the corresponding [PJN], whose projected trajectory can be of different size, from the immediate local to the further global one (cf. [14]; [15]).

3. Methodology

In the following, we first describe the methodology for calculating the emphasis density throughout the trajectory of both types of information units. In section 3.2 we present the method for estimating the location of the information planning units by their trajectory size within BG and PG.

3.1. Emphasis density

To calculate the emphasis density, our first step involves merging the reduction E0 label with E1 in the perceived emphasis annotations for the spontaneous speech data, i.e. SpnL and SpnC. Since read speech data are without the reduction annotation, we merge the E0/E1 labels so as to provide a consistent standard for the further calculation of emphasis density across different speech genres. For the scoring of the emphasis density, we then assign *arbitrarily* a 0 score to the ET label E1, and labels E2 and E3 receive the incremental scores of 1 and 2 respectively.

Following (1), we calculate **emphasis density** (ED) by the **PW** unit, while taking into consideration the emphasis levels annotated for the pre-/post- PWs to the current PW. In the end, the average emphasis density score is derived from not merely each individual PW, but together with averaged scores from PWs in its neighborhood.

$$ED = Ave(\text{Pre_PW_Score} + \text{Cur_PW_Score} + \text{Post_PW_Score}) \quad (1)$$

3.2. Locating information planning units at higher-level DPUs

To test the hypothesis regarding the size of the information units in relation to their location within higher level discourse prosodic units, we further examine the location of the two types of information planning units at BG and PG levels. Given that BG and PG can vary drastically across different speech genres (cf. [16]), the first step involves a normalization of BG- and PG-level units. Then we calculate the averaged position of the starting point of [PJR], the ending point of [PJN], as well as that of [KEY], all within the normalized BG/PG units. It is further noted that since the [PJR]-[PJN] unit can be realized in different trajectory size, we present the results of average positions of the [PJR]-[PJN] unit by PPh.

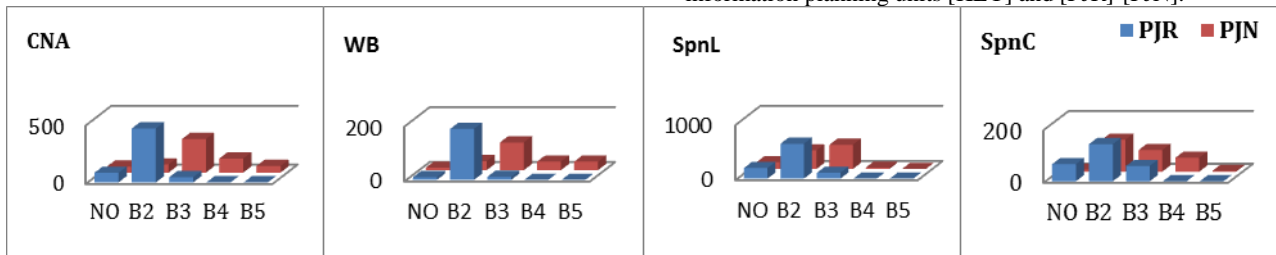


Figure 1: The distribution of [PJR] and [PJN] boundary endings by DPU levels.

4. Results and discussions

4.1. The information units and DPU boundaries

As explained in section 2.2.3, the categorization of information content has been based on PW. On the other hand, the corresponding [PJN] to each annotated [PJR] can be realized in various trajectory size and of different trajectory scope. Here we start out by providing evidences explaining why together [PJR] and [PJN] should be considered as a unit. Specifically, we examine the corresponding DPU boundaries at the end of both [PJR] and [PJN]. The results of boundary distribution are presented in Fig. 1, and Table 2 summarizes the relative [PJR]/ [PJN] ending boundaries (i.e. if the [PJR] ending boundary is of lower/higher DPU levels or equal to [PJN] ending boundary level) in proportion.

Table 2: The relative [PJR] and [PJN] ending boundaries in comparison.

Genre Boundary ending	CNA	WB	SpnL	SpnC
PJR<PJN	78%	81%	52%	49%
PJR=PJN	13%	15%	35%	40%
PJR>PJN	8%	4%	13%	14%

4.1.1. Discussion

From Fig. 1, we find an obvious tendency across the 4 speech genres in that, for most cases the endings of [PJR] fall at the DPU boundary of *prosodic word* (PW). The [PJN] endings, on the other side, tend to fall at the boundary of *prosodic phrase* (PPh) unit or even higher level discourse units of BG or PG. The results indicate that [PJR] is more of a prosodic prompt phenomenon at the local word-level, while [PJN] can extend from the word-level upward and to the higher discourse levels. Further supporting evidences can be derived from Table 2, which shows that about 80% of the read speech data and at least 50% of the spontaneous data have their [PJN] ending boundaries falling by DPUs higher than the [PJR] ending boundaries.

The findings, therefore, provide solid evidence to illustrate why the prosodic prompted [PJR] should be considered together with its respective [PJN] as a unit in the process of information planning. Moreover, since the trajectory of projection could be of different size, we further examine the range distribution of the [PJR]-[PJN] pair by the unit of PPh. The result is summarized in Table 3 (next page). Based on the observation, we take PPh as the base unit of further analyses in the following discussions towards the allocation and compensation between the two types of information planning units [KEY] and [PJR]-[PJN].

Table 3: Trajectory size distribution of [PJR]-[PJN] information unit by numbers of PPhs.

Genre PPh #	CNA	WB	SpnL	SpnC
1	63%	77%	66%	55%
2	25%	13%	18%	28%
3	6%	3%	7%	8%
Over 3	6%	7%	9%	9%

4.2. Calculation of emphasis density

Following the methodology described in 3.1, we first carry out the calculation of the emphasis density. As illustrated previously in Table 3, [PJR]-[PJN] as an information unit can mostly be accounted for by at least 1 PPh and over 90% can be accounted for by up to 3 PPhs. Therefore in the following analysis we take **PPh** as the base unit and calculate the emphasis density by either [KEY] or [PJR] annotated within a PPh. The results are summarized in Figures 2 and 3.

4.2.1. Discussion

As shown, from Figures 2 and 3 a general tendency of **heavy-to-light** emphasis density can be observed over the trajectory of both [KEY] as well as [PJR]-[PJN] units. The only exception is with the speech data of WB (i.e. the 2nd panels of Figures 2 and 3), in which a heavier emphasis density is found towards the end of [KEY] unit, especially when extending over 2 PWs, also towards the end of the trajectory when the unit extends as long as 6 PWs. The heavy-to-light tendency across the two types of information units otherwise reflects that speakers almost always start out by planning the heaviest information density and lightened up the information loading all the way through the trajectory of both unit types. As for the exception found in WB data, it is suggested perhaps this may be due to the more evenly distributed focal information across the information units in this specific speech genre.

Another interesting point to address is that, based on Figures 2 and 3, an obvious speech-genre distinction can be observed: whereas the read speech CNA (i.e. the 1st panels of Fig. 2 and 3) demonstrates a more gradual slope from heavy to light across the trajectory of the information units, the spontaneous speech data SpnL and SpnC show a much

steeper slope from the beginning of the units. Most of all, among the spontaneous speech data, it is the conversation data SpnC that exhibit the steepest slope of all, starting from the beginning of the trajectory through the end. It is assumed that this may be contributed from that the spontaneous speech data have been annotated with *reduction*, which in turn lowers drastically the overall scoring of emphasis density. This is especially significant with the spontaneous conversation, which otherwise reflects speakers constantly and extemporaneously adjusting their speech production by incorporating and inserting in elements such as fillers and discourse markers.

4.3. Correlation between information units and their positions at BG

In this section we turn to the hypothesis regarding the correlation between the information units and their positions in terms of the DPU of BG. Following the methodology described in section 3.2, we observe the averaged positions of the [PJR] starting point and the [PJN] ending point, as well as [KEY], by the trajectory size of the units. Note here the base discourse-prosodic unit for comparison is PPh. The results are summarized in Fig. 4 (next page, top panels).

4.3.1. Discussion

From Fig. 4 it is demonstrated that when the [PJR]-[PJN] unit becomes longer (i.e. expands over 2 PPhs), the averaged starting point of its [PJR] would shift towards the beginning of the normalized BG. At the same time, the longer the [PJR]-[PJN] pair, the further the trajectory of the projection would be, and hence the corresponding [PJN] would fall towards the end of the normalized BG, thus forming a coherent head-tail echo. Interestingly, the result of the information unit [KEY] reflects that in average it is located towards the center of the normalized BG. Of course such result is reflected from the fact that in the original annotation [KEY] has been based on PW. Nevertheless, this finding of averaged position does reinforce the fact that prosodic prompted [KEY] (key information) is distributed evenly across the entire BG.

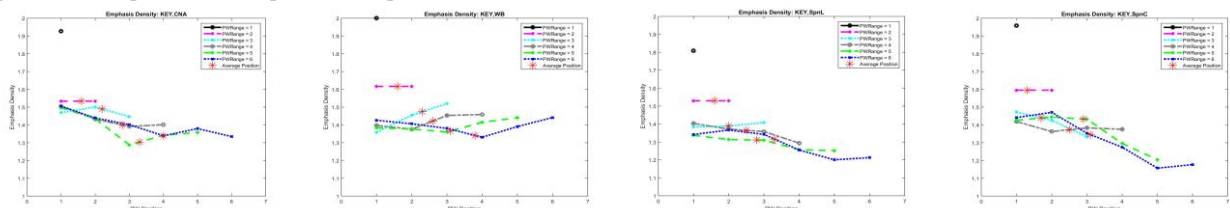


Figure 2: Results of emphasis density from [KEY] information planning units across speech genres (by PW).

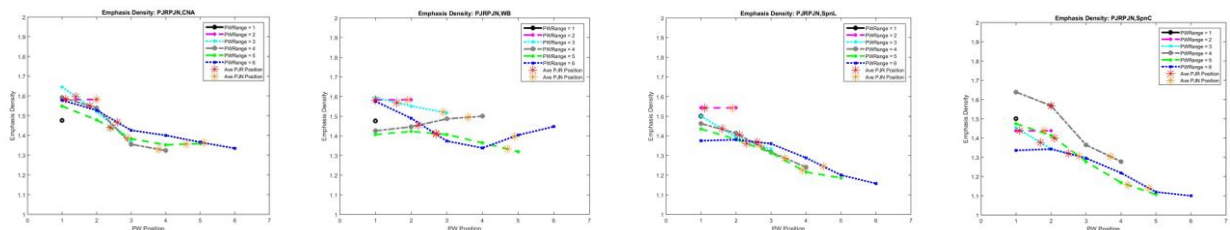


Figure 3: Results of emphasis density from [PJR]-[PJN] information planning units across speech genres (by PW).

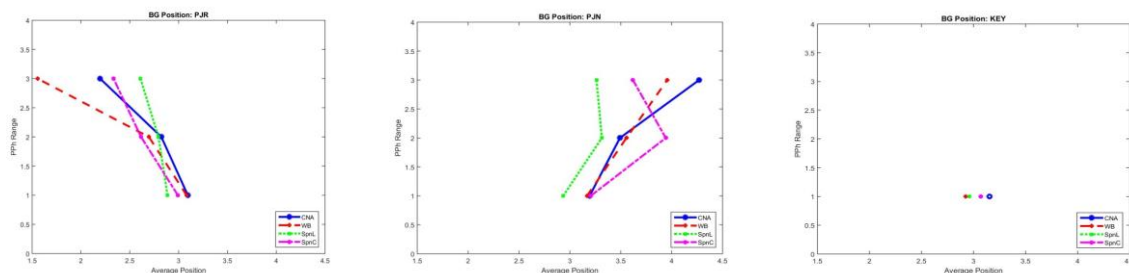


Figure 4: The correlation between the size of information units and their locations within normalized BG.

The results, therefore, confirm the proposed hypothesis that a positive correlation can be identified between the location of the information unit and its relative size, especially by [PJR]-[PJN]. In other words, for the planning of prosodic prompted [PJR] throughout its projection trajectory, the size of the *projection* in planning is a crucial factor. For planning a larger projection trajectory from the [PJR]-[PJN] unit, the speaker would have to estimate and arrange for an earlier start in order to accommodate the projected completion of the entire unit. Remarkably, the planning of the [PJR]-[PJN] information unit forms an interesting *compensatory* relationship with the planning of prosodic prompted [KEY]: speakers may plan and signal the [KEY] information via prominence cues whenever there is a need to do so; it can be located at any possible position within the entire BG.

4.4. Correlation between information units and their positions at PG

The above results lead us to ponder if the same findings will hold when turning to the largest discourse-prosodic unit in planning of the continuous speech, namely the *multiple phrase speech paragraph* PG. Therefore in this section we turn to the correlation between the information units and their positions within the normalized PG. The results are summarized in Fig. 5.

4.4.1. Discussion

Results from Fig. 5 indicate that, the location of the [PJR] starting point within the normalized PG does not reflect as distinctively different as the size of the projection trajectory changes, such as when the unit expands over 2 PPhs. Moreover, the comparison between the position of the [PJR] starting point and [PJN] ending point demonstrates quite similar locations in terms of the whole PG, i.e. the averaged positions of where the information unit starts and ends are fairly close to each other and centered within the normalized PG. Meanwhile, the averaged location of the [KEY] unit is also placed fairly centered to the normalized PG, similar to

what has been identified regarding its location in BG. From these results, they suggest that both types of information units demonstrate an even distribution across the normalized PG.

Given the findings on the relative positions of the two types of information units within BG and PG, therefore, it is further suggested that the execution of planning for the information units is most likely carried out based on the DPU of BG. In other words, the allocation and planning of information at the higher level discourse units is organized most significantly by BG. Specifically, whereas speakers plan for the deployment of information by the discourse-prosodic unit of BG, it is most likely that the top level PG is the unit for the topic-based discourse planning and processing.

5. Summary and conclusions

The current research explores two major types of prosodic prompted information units *key information* [KEY] and *projector-projection* [PJR]-[PJN] pair in continuous speech, building upon the assumption that perceived prosodic highlight allocations is directly associated with the information planning. Although prosodic highlight has been traditionally linked to the indexing of focal and salient information, it is held here that another far more significant function of prosodic prominence is the speaker's advanced cuing of the soon-to-arrive key information throughout the planned trajectory of *projection*, thereby prompting expectation online of the up-coming focal information.

The examination of allocations of the two types of information units has been carried out, first of all, via the calculation of *emphasis density* by either [KEY] or [PJR] annotated within a PPh. The results demonstrate a consistent heavy-to-light deployment of emphasis density, and hence a decreasing of information loading from the starting to the ending of the trajectory of both units. The constant pattern identified otherwise reinforced that speakers tend to plan for the heaviest information loading from the beginning of both units. Most of all, this finding echoes the suggestion from [5]

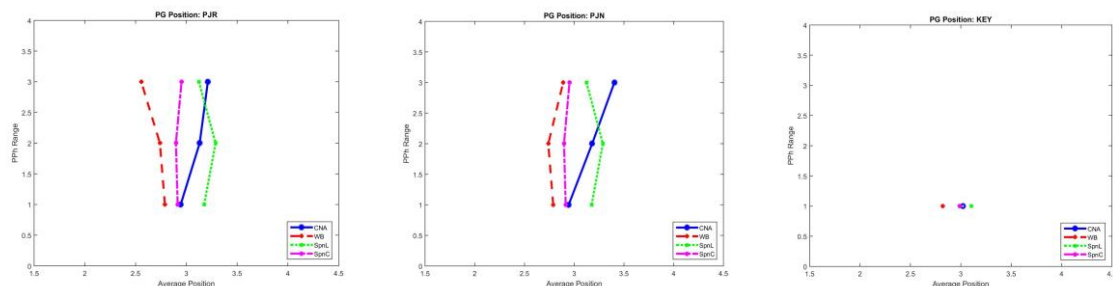


Figure 5: The correlation between the size of information units and their locations within normalized PG.

that 'speakers face a phase of high cognitive stress in the beginning of (syntactic) projections'. Here supportive evidence to the claim about the heaviest loading of cognitive stress at projections beginning is further supported by examining the allocation of emphasis density across the major information planning units.

We further test the hypothesis regarding the correlation between the location of both types of information units and their trajectory size. By the normalized discourse-prosodic level of BG, it is identified that for the [PJR]-[PJN] unit, the locations of the starting and ending points are positively correlated with its trajectory size: the larger the projection trajectory, the earlier its beginning and the later its ending would be located within the normalized BG. On the other hand, the analysis indicates that [KEY] as a basic information unit can be distributed evenly across BG. As result, the placement of both information units within BG not only indicates a *compensatory* relationship in terms of their locations, but also signifies the most crucial aspect of the advanced prosodic prompting. Since we did not find similar results at the normalized PG level, it is thus suggested that the information planning is mostly carried out at the BG level.

In sum, the current study contributes to an inclusive understanding towards the mechanism behind information planning in continuous speech, specifically the allocation and the compensation between the two types of prosodic highlight prompted information units. We therefore propose that the [KEY] and [PJR]-[PJN] pair form the two major units in information planning for continuous speech while the new-vs.-old information status can only offer part of the explanation to the comprehensive information structure. In addition, we would like to stress the significance of advanced prompting and projection of information planning in continuous speech. For future research, we plan to substantiate the current findings through more empirical tests and perception experiments based on continuous speech data to further validate the functions and relationship identified for the prosodic highlight in continuous speech.

6. References

- [1] Falk, S. 2014. On the notion of salience in spoken discourse-prominence cues shaping discourse structure and comprehension. *Travaux interdisciplinaires sur la parole et le langage*, (30).
- [2] Halliday, M. A. K. 1967. Notes on transitivity and theme in English. *Journal of Linguistics* 3, 199-244.
- [3] Pierrehumbert, J. B.; Hirschberg, J. 1990. The meaning of intonational contours in the interpretation of discourse. In *Intentions in communication*, P. Cohen et al. (eds.), Cambridge: MIT Press, pp. 271-311.
- [4] Chen, H.; Fang, W.; Tseng, C. 2016. Prosodic prompts of information content in speech - A cross genre comparison of prominence as key, projector and projections. The 24th Annual Conference of International Association of Chinese Linguistics, July 17-19, Beijing.
- [5] Auer, P. 2009. On-line syntax: thoughts on the temporality of spoken language. *Language Sciences*, 31:1-13.
- [6] De Ruiter, J; Mitterer, H; Enfield, N. J. 2006. Projecting the end of a speaker's turn: a cognitive cornerstone of conversation. *Language*, 82 (3), 515-535.
- [7] Chomsky, N. 1986. *Knowledge of language: Its nature, origin, and use*, Greenwood Publishing Group.
- [8] Haegeman, L. 1994. *Introduction to Government and Binding Theory*, Blackwell Publishing.
- [9] Schegloff, E. A. 1996. Turn organization: One intersection of grammar and interaction. In *Interaction and grammar*. Cambridge: Cambridge University Press, 52-133.
- [10] Ford, C. E.; Thompson, S. 1996. Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. In *Interaction and grammar*, Cambridge: Cambridge University Press, 135-184.
- [11] Pin, S.; Lee, Y.; Wang, H.; Chen, Y. 2008. Fluent speech prosody: Framework and modeling. *Speech communication*. 46 (3-4), 284-309.
- [12] Tseng, C.; Su, Z. 2008. Discourse prosody and context-global F0 and tempo modulations. *INTERSPEECH 2008 - 9th Annual Conference of the International Speech Communication Association*, Brisbane, Australia, 1200-1203.
- [13] Tseng, C. 2010. An F0 analysis of discourse construction and global information in realized narrative prosody. *Language and Linguistics* (11)2, 183-218.
- [14] Lerner, G. 1996. On the 'semi-permeable' character of grammatical units in conversation: Conditional entry into the turn space of another speaker. In *Interaction and grammar*. Cambridge: Cambridge University Press, 238-271.
- [15] Huang, S. 2013. Projection and repair. In *Chinese grammar at work*, S. Huang (ed.). Amsterdam: John Benjamins, 321-346.
- [16] Chen, H.; Fang, W.; Tseng, C. 2015. Information content, weighting and distribution in continuous speech prosody - A cross-genre comparison. *Oriental-COCCOSDA 2015*. Shanghai, China, 75-80.