Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

The role of prosodic cues in the perception and anticipation of sentence completion in structurally ambiguous sentences

Yujie Ji[1], *Xiaoming Jiang*[1, 2]

[1]Institute of Linguistics, Shanghai International Studies University
[2]Key Laboratory of Language Sciences and Multilingual Artificial Intelligence, Shanghai International Studies University
yujie_ji@shisu.edu.cn, xiaoming.jiang@shisu.edu.cn

Abstract

In spoken discourse, speakers strategically deploy prosodic cues to shape listeners' perceptions and anticipations of sentence completion to guide their interpretation of structurally ambiguous sentences. By asking 64 listeners to judge the perceived sentence completeness on a 11-pt scale (Exp. 1) or to anticipate the ending point of the sentence by pressing a button (Exp. 2), this perceptual-acoustics study explored three aspects: (1) Acoustic distinctions at various levels of prosodic hierarchy between different ambiguity types; (2) The influence of prosodic cues at different levels on the perceived sentence completeness; (3) Which level of prosodic cues was employed by the listener to perceive and anticipate sentence completion. The findings revealed that: (1) Structural ambiguity types impacted f0, intensity, and duration of boundary nouns; (2) Compared with sentences with structural ambiguity, local prosodic cues of sentences without ambiguity increased perceived completeness and reduced RT for anticipation for the sentence-completion, while distal cues showed no significant impact; (3) Compared to prosodic cues at a single level, the predictive ability of prosodic cues at multiple levels was significantly stronger. These outcomes suggest that listeners rely on local acoustic parameters to determine spoken sentence completion to guide interpretation of structural ambiguity, supporting local encoding account, whereas speakers take a "global" mechanism to encode different interpretations of structural ambiguity. These findings underscore the interplay between listener perception and speaker encoding in establishing perceived prosodic boundaries in spoken language.

Index Terms: Prosodic hierarchy; Chinese syntactic ambiguity; Speech perception; Perception and anticipation of sentence completeness

1. Introduction

In oral communication, how do speakers encode different interpretations of syntactic structures through the verbal information in their speech? How do listeners perceive prosodic differences between different interpretations of ambiguous syntactic structures through the encoded vocal information in order to facilitate sentence comprehension?

For different communicative purposes, speakers may encode different information through similar verbal structures[1-3]. When individuals express requests to others in Chinese, they use simple structures to indicate the target; or clarify detailed information about the target by adding lexical clues; or specify the details of the object with a temporary local semantic mismatch.

Although speakers used the same phrase structure V+PP+NP in all three contexts, their understandings of whether the structure ends vary. During an oral conversation, speakers encoded specific prosodic cues, known as prosodic boundaries which are prosodic features of discourse boundaries in oral communication that signal the completion of sentences. Different levels of prosodic boundaries facilitated listeners' comprehension of conversation, with their acoustic and location features affecting listeners' comprehension, such as facilitating the resolution of temporary structural ambiguities[4] [5], and facilitating the fluency of a speaker's oral expression [6][7]. Prosodic boundaries were perceived by both communicators in spoken conversations based on the acoustic cues[8], such as (1) pitch changes in prosodic boundary; (2) duration of word before prosodic boundaries; and (3) pauses [1] [9].

Previous studies investigating the relationship between prosodic boundaries and syntactic hierarchy found that the two influenced each other. On one hand, syntactic structure affected the perception of prosodic boundaries. Listeners' anticipations of syntactic endpoints affected the perception of intonation boundaries[10]. Regardless of the levels of prosodic boundaries, listeners' judgments reported more prosodic boundaries at syntactically permissive place than where is impermissible. This suggests that listeners' judgments of prosodic boundaries were not strictly driven by acoustic cues, but were influenced by syntactic cues. Well-structured syntax facilitated the perception of prosodic boundaries[11]. On the other hand, prosodic boundaries also influenced the comprehension of spoken sentence structure. Behavioral and EEG studies found that prosodic boundaries signal the syntactic structure of a sentence [11-13], suggesting that prosodic boundaries were acoustic representations of syntactic structure at the phonological level. In addition, prosody cues influenced anticipations of spoken sentence turn endings in conversation[14]. Listeners use prosodic boundaries cues to decode the meaning of sentence endings. Furthermore, when processing double prepositional phrase disambiguation structures, according to the principle of *late closure*, the second prepositional phrase in a double prepositional phrase structure should modify the nearest noun, whereas in fact, if a prosodic boundary precedes the second prepositional phrase, listeners tended to interpret the second preposition as the closure of the sentence if there was a prosodic boundary before the second preposition. Prosodic cues were as central to sentence processing as the syntactic hierarchy[6] [15], and prosodic cues did not merely play a supportive role in sentence comprehension.

Clearly, prosodic boundaries play a crucial role in introducing deviations or resolving ambiguity in structural interpretation. While prior studies have explored the acoustic

 10.21437/SpeechProsody.2024-244

representation of prosodic boundaries in different structural contexts and their impact on sentence comprehension, they fall short in revealing the acoustic encoding and perceptual mechanisms specifically in temporarily ambiguous structures. Moreover, a debate exists on whether processing prosodic boundary information relies on local or global prosodic cues. The local processing account suggests[16][17] that prosodic boundary is determined by cues closely tied to the boundary, influencing comprehension locally and context-independently. Conversely, the global processing account argues that the strength of prosodic boundary features in adjacent hierarchies affects prosodic boundary perception (Informative Boundary Hypothesis [18]). Evidence indicates that speakers encode information early in sentences [19], demonstrating an interplay between prosody and structure not only at syntactic boundaries but also at the beginning of sentences.

Hence, this study leverages the syntactic characteristics of the Chinese language, specifically focusing on the ambiguous structures within Chinese object complement clauses[20]. All sentences were equal in number of characters. Three distinct sentence structures are created as critical conditions, which are:

(1) Syntactic Boundary Disambiguation (S+P+: Complete syntactic structure, complete prosodic boundary): 递给我那瓶水 ([V 递: pass][PP 给我: to me][nP[D 那瓶: that bottle of [NP 水: water]]])

(2) Syntactic Category Disambiguation (S+P-: Complete syntactic structure, incomplete prosodic boundary): 递给我那瓶水边上笔 ([V 递: pass][PP 给我: to me][nP[NP 那瓶水 that bottle of][n 边上的: beside[NP 笔 :pen]])

(3) Semantic Violation Disambiguation (S-P-: Incomplete syntactic structure, incomplete prosodic boundary): 递给我那支水边上的笔 ([V 递: pass][PP 给我: to me][nP[n 那支: that[NP[n 水边上的 beseide the water [NP 笔: pen]])

In the three sets of sentences, the temporarily ambiguous structures at the first syntactic completion noun (e.g., "水") can all serve as potential sentence endpoints. However, different cues within the sentences alter the ultimate interpretation of this structure. For the speaker, the distinction between (1) and (2) lies in whether there is an explicit syntactic boundary after "水" in (1) or if additional syntactic cues are provided, thus altering the syntactic category of "水" in (2). The difference between (2) and (3) hinges on the presence of a semantically mismatching phrase (e.g., 那支水), facilitating the reinterpretation of the ambiguous structure in (3). Employing syntactic or semantic cues guides speakers to produce temporarily ambiguous phrases with different interpretations, aiding in the encoding of prosodic boundary features based on distinct structural understandings. Speakers provide effective prosodic disambiguation information only when they are aware of explicit structural ambiguity requiring resolution through prosodic cues [21].

The present study aims to address three questions: first, does differential acoustic representation exist between prosodic boundary cues at different levels for various interpretations of ambiguous phrases; Based on previous research, this study hypothesizes that prosodic boundaries of different structures may be characterized by differences in acoustic parameters such as f0, pause duration, and intensity. Second, are there perceptual differences in the perception and anticipation of sentence-completion of identical ambiguous phrase structures; third, how does prosodic boundary information influence listeners' passive perception and active anticipation of completion of different phrase structures. According to the local processing account, only the prosodic boundary at the sentence-final word will influence the listener's judgment of sentence ending, whereas according to the global processing account, prosodic boundaries throughout the entire sentence can contribute to the listener's judgment of sentence ending.

2. Method

2.1. Materials

A total of 36 sets of sentences were created. To ensure the probability of the two nouns as the role of object in the sentence was similar, an additional 36 sets of sentences with reversed version of nouns were designed (e.g., 递给我那瓶水边上的笔 vs. 递给我那支水边上的笔), resulting in a total of 72 sets.

Two native Chinese speakers (1 male, 1 female, both aged 19) were recruited for pronunciation. The male speaker had 13 years of education, 14 years of musical training, and 1 year of drama performance training. The female speaker had 13 years of education, 4 years of musical training, and 1 year of vocal training. Recordings were conducted in the soundproof room of the Speech Lab at Shanghai International Studies University, using an Audio-Technica AT2032 microphone, NI Komplete Audio 6 MK2 sound card, and recorded at a 44.1 kHz sampling rate with Praat software.

During the production experiment, speakers were instructed to understand the meaning of each sentence first and ensure that listeners could comprehend their intended meaning through their speaking style. Each sentence was uttered twice. After recording, S+P- and S-P- sentences were edited from the beginning to the first completion word (递给我那瓶水 or 递给我那只水), so that the edited sentences and S+P+ as its entirety were presented.

2.2. Experiments: Perception and Anticipation of Sentence Completion

Participants reported to have no language or speech disorders, no hearing impairments, no neurological disorders, and no mental disorders. All participants provided informed consent. The perception experiment was approved by the Ethics Committee in the Institute of Linguistics at Shanghai International Studies University.

2.2.1. Experiment 1: Perception of Sentence Completion

Thirty-two native Mandarin speakers (14 males, 18 females, M=20.47 years, *SD*=2.78 years) were recruited for the perception experiment. Each participant was asked to rate the sense of completion for each auditorily presented sentence on a scale of 0 to 10 (0 indicating that the sentence definitely was not finished and there was more content to follow; 10 indicating that the sentence was definitely complete, and there was no possibility to continue the sentence).

2.2.2. Experiment 2: Anticipation of Sentence Completion

An independent group of 32 native Mandarin speakers (16 males, 16 females, M=20.24 years, *SD*=3.14 years) participated in the anticipation experiment. Each participant was instructed to press a key with their right hand immediately upon perceiving that the speaker was about to end the utterance. The instruction emphasized not waiting until the speaker finished asking the question and ceased speaking before responding. Taking into account the anticipated influence of mismatches between S-P-phrase quantity names, the expected experimental stimuli in this section exclusively comprised of S+P+ sentences and S+P- sentences.

2.3. Data Analysis

The Prosody Pro script in Praat was used for extracting acoustic parameters, including pause duration before the prosodic boundary word, word duration, intensity (mean intensity), and fundamental frequency (mean f0).

To examine the acoustic representation underlying the perceived sentence completeness and the anticipation of sentence ending under different syntactic conditions, a mixed linear model was established for each acoustic parameter (pause duration, word duration, mean intensity, mean f0) with syntactic structure as the independent variable and speaker, sentence, version, and repetition as random variables.

To investigate how prosodic boundary information influences listeners' perception of sentence completeness, multiple linear regression models were applied. Sentence completion scores were used as the dependent variable, and the acoustic parameters of the prosodic boundary words (duration, mean f0, and mean intensity) served as predictor variables.

To explore how global or local prosodic boundary information influences the completeness perception, three linear regression models were built. Sentence completion scores were used as the dependent variable, and acoustic parameters at different levels of the prosodic boundary, specifically, those at the boundary word level ("水"), the prosodic phrase level (including the boundary word and two proceeding characters "那瓶" and), and the whole sentence level (including the initial two characters "拿给我", adjacent components and the boundary word), served as predictor variables. The predictors at each level entered the model in a sequential way (see below).

The model selection process started with the Boundary word level model, including only the acoustic parameters of the local boundary word. Additional, more global, prosodic predictors were then added stepwise to determine model fit, using the likelihood ratio test based on the models' AIC (Akaike Information Criterion). Predictors that did not significantly improve the model fit were removed from the models [22]. The analysis was conducted using the linear mixed-effects model packages (lmerTest) in the R language.

3. Results

3.1. Prosodic Representation of Different Levels of Boundaries in Ambiguous Phrases

3.1.1. Prosody of Boundary Words

Our analysis of pause duration near word boundaries revealed significant distinctions between different structures (S+P+, S+P-, S-P-; $F(2,824)=111.93$, $p<.001$). Specifically, S+P+ exhibited a notably shorter pause duration compared to S-P- (51.51 ms vs. 99.93 ms, $\beta=-55.6$, $t=-13.65$, $p<.001$), while S+P- displayed a shorter pause duration than S-P- (60.48 ms vs. 99.93 ms, $\beta=-49.4$, $t=-12.13$, $p<.001$). However, there was no significant difference in pause duration between S+P+ and S+P- ($\beta=-6.2$, $t=-1.52$, $p=0.386$).

Additionally, the analysis of word duration at boundaries showed significant differences across structures ($F(2,825)=8.25$, $p<.001$). Specifically, S+P+ had a longer word duration than S+P- (376.08 ms vs. 306.85 ms, $\beta=24.5$, $t=3.23$, $p=.004$), while S-P- and S+P+ exhibited no significant difference in word duration (311.33 ms vs. 376.08 ms, $\beta=-3.9$, $t=-0.52$, $p=1.000$). Notably, S-P- had a significantly longer word duration than S+P- ($\beta=-28.4$, $t=-3.75$, $p<.001$).

There were no significant differences in the mean f0 of boundary words across the different sentence structures ($F(2,814)=0.88$, $p=.415$).

Analysis of the mean intensity of boundary words revealed significant differences across structures ($F(2,825)=4.67$, $p=.010$). Specifically, in S+P+, the boundary words exhibited significantly lower intensity than those in S+P- (41.47 dB vs. 44.00 dB, $\beta=-0.61$, $t=-2.54$, $p=.034$) and S-P- (41.47 dB vs. 44.39 dB, $\beta=-0.66$, $t=-2.74$, $p=.019$). However, there was no significant difference in intensity between S+P- and S-P- ($\beta=0.05$, $t=0.20$, $p=1.000$).

3.1.2. Prosody of Adjacent Boundary Word Components

Our investigation into the duration of adjacent boundary word components revealed significant differences across structures ($F(2,824)=252.38$, $p<.001$). Specifically, in S+P+, the components had significantly longer duration than those in S+P- (734.82 ms vs. 670.86 ms, $\beta=64.0$, $t=-7.85$, $p<.001$). Additionally, S+P+ had significantly shorter component duration than S-P- (734.82 ms vs. 851.29 ms, $\beta=-116.0$, $t=-14.30$, $p<.001$), and S+P- had significantly shorter component duration than S-P- ($\beta=-180.0$, $t=-22.16$, $p<.001$) .

In terms of the mean f0 of adjacent components, our results showed significant differences between structures ($F(2,824)=53.68$, $p<.001$). Specifically, the f0 in S+P+ was significantly lower than in S+P- (172.07 ms vs. 181.09 ms, $\beta=-9.02$, $t=-8.92$, $p<.001$) and S-P- (172.07 ms vs. 181.19 ms, $\beta=-9.12$, $t=-9.02$, $p<.001$), while there was no significant difference between S+P- and S-P- ($\beta=-.10$, $t=-0.57$, $p=1.000$).

Moreover, the analysis of the mean intensity of adjacent boundary word components indicated significant differences between structures ($F(2,824)=130.65$, $p<.001$). In S+P+, the components had significantly lower intensity than those in S+P- (42.61 Hz vs. 44.15 Hz, $\beta=-1.54$, $t=-12.97$, $p<.001$) and S-P- (42.61 Hz vs. 44.38 Hz, $\beta=-1.76$, $t=-14.84$, $p<.001$). However, there was no significant difference in intensity between S+P- and S-P- ($\beta=-0.22$, $t=-1.87$, $p=.184$).

3.1.3. Prosody of Initial Component Boundaries

The analysis of the duration of initial components indicated significant differences between structures ($F(2,824)=17.83$, $p<.001$). Specifically, the initial components in S+P+ had a significantly longer duration than those in S+P- (436.56 ms vs. 418.48 ms, $\beta=18.08$, $t=4.08$, $p<.001$), while the components in S+P+ were significantly longer than those in S-P- (436.56 ms vs. 410.76 ms, $\beta=25.80$, $t=5.82$, $p<.001$). Notably, the components in S+P- were significantly longer than those in S-P- ($\beta=7.72$, $t=1.74$, $p=.246$).

Additionally, the analysis of the mean f0 of initial components showed significant differences between structures ($F(2,824)=20.13$, $p<.001$). The f0 in S+P+ was significantly lower than in S+P- (202.04 Hz vs. 208.71 Hz, $\beta=-6.67$, $t=-6.28$, $p<.001$) and S-P- (202.04 Hz vs. 204.51 Hz, $\beta=-2.47$, $t=-2.32$, $p<.001$), while the f0 in S+P- was significantly higher than in S-P- ($\beta=4.20$, $t=3.95$, $p<.001$).

Furthermore, the analysis of the mean intensity of initial components revealed significant differences between structures ($F(2,824)=27.76$, $p<.001$). The intensity of the initial components in S+P+ was significantly lower than in S+P- (46.69 dB vs. 47.40 dB, $\beta=-0.70$, $t=-5.33$, $p<.001$) and S-P- (46.69 dB vs. 47.63 dB, $\beta=-0.94$, $t=-7.17$, $p<.001$). However, there was no significant difference in intensity between S+P- and S-P- ($\beta=-0.24$, $t=-1.84$, $p=.197$).

3.2. Differences in Sentence Completeness and RTs for Expecting Sentence Ending between Ambiguous Phrases

The average perceived completeness score was 8.26 (*SD*=2.33) for S+P+ sentences, 5.19 (*SD*=3.65) for S+P- sentences, and 2.43 (*SD*=3.24) for S-P- sentences. One-way ANOVA indicated a significant difference in perceived sentence completeness among different sentence cues (F=(2,27557)=8185.1, p<.001). Specifically, the completeness score for S+P+ sentences was significantly higher than that for S+P- sentences (β=3.07, t=67.38, p<.001), and S+P+ sentence completeness score was significantly higher than that for S-P- sentences (β=5.84, t=127.89, p<.001). Additionally, the completeness score for S+P- sentences was significantly higher than that for S-P- sentences (β=2.76, t=60.50, p<.001).

Concerning RT for expecting sentence ending, there were significant differences between sentence types with varying prosodic cues (F=(1,8484.6)=605.0, p<.001). The average RT for S+P+ sentences (M=0.32, SD=0.43) was significantly faster than that for S+P- sentences (M=0.57, SD=0.61).

3.3. Prosodic Cues Predicting Recognition of Sentence Completion

To investigate how prosodic boundary information influences listeners' perception of sentence completion, this analysis employed multiple linear regression. The dependent variable was the perceived sentence-completeness score, while the predictors were duration, mean f0, and mean intensity of the prosodic boundary word. The model demonstrated statistical significance (F(5,26315)=543.3, p<.001), with the prosodic boundary word's duration, mean f0, and intensity accounting for 9.34% of the variance in the perceived sentence-completeness scores (adjusted R^2 = 0.093).

Furthermore, to explore whether global or local prosodic boundary information impacts such recognition, another two hierarchical models were constructed using predictors from the prosodic boundary word level, the prosodic boundary phrase level, and the entire sentence level.

Model 2, representing the prosodic boundary phrase level, demonstrated statistical significance (F(8,26311)=549.5, p<.001), explaining 14.29% of the variance in sentence-completeness perception scores using acoustic parameters from both the prosodic boundary word and its adjacent word components (adjusted R^2 = 0.143). A comparison between Model 1 and Model 2 revealed a significantly better fit for the phrase level model (F(3,26314)=950.24, p<.001).

Adding sentence-level prosodic cues to the model resulted in Model 3, which was statistically significant (F(11,26309)=429.8, p<.001). This model explained 15.2% of the variance in sentence-completeness scores by incorporating acoustic parameters from the prosodic boundary word, adjacent components, and the sentence's initial component (adjusted R^2 = 0.152). Comparing Models 2 and 3 revealed that the sentence-level prosodic boundary model had a significantly better fit (F(3,20178)=507.53, p<.001).

3.4. Prosodic Cues Predicting Anticipation of Sentence Ending

To investigate how prosodic boundary information influences listeners' proactive anticipation of sentence completion, this study again formulated three linear regression models. The dependent variable was the reaction time associated with anticipation of sentence completeness. The findings revealed statistical significance of Model 1 (F(3,8476)=19.88, p<.001), indicating that 0.7% of the variance in RT was explicated by variations in the acoustic parameters of boundary word (adjusted R^2 = 0.007).

At the prosodic boundary phrase level (Model 2), the multiple linear regression model demonstrated statistical significance (F(6,8476)=36.17, p<.001), indicating that 2.42% of the variance in RT can be clarified by acoustic parameters in both boundary word and adjacent word components (adjusted R^2 = 0.024). A comparative analysis between Model 1 and Model 2 underscores a significantly superior fit for the latter (F(3,8473)=52.11, p<.001).

Adding sentence-level prosodic cues resulted that Model 3 was also significant (F(9,8476)=25.37, p<.001). This model indicates that 2.52% of the variance in perceived sentence completeness can be clarified by variations in the acoustic parameters of boundary word, adjacent word components, and sentence initial component (adjusted R^2 = 0.025). A model comparison between the whole-sentence prosody Model 2 and Model 3 indicated no significant difference in the goodness of fit (F(3,8473)=3.71, p=.011).

4. Discussion and Conclusion

This study initially investigates how speakers utilize prosodic boundaries to encode acoustic features of temporarily ambiguous structures (Chinese object complement structures). The prosodic boundaries with stronger ending point perception display longer duration, shorter pause duration, and lower intensity compared to those with weaker ending point perception. Additionally, this study discovers that prosodic cues at different hierarchical levels influence the perception of syntactic structures. In comparison to those of a single level, the predictive ability of prosodic cues at multiple levels becomes stronger.

Previous researches suggest that acoustic representations of different hierarchical prosodic boundaries vary, and features such as duration of consonants at the end of sentence constituents, pauses, or pitch serve as reliable prosodic cues[8][23][24]. This ultimately results in different acoustic representations of prosodic boundary words under distinct structural types, indicating that speakers can use prosodic cues to encode the acoustic features of phrase structures in different communicative contexts.

Furthermore, from the listener's perspective, prosodic cues impact the perception of the degree of finality in ambiguous structures. The duration of boundary words, intensity, and pauses serve as cues utilized by listeners to assist in understanding whether a sentence has completed. Consistent with previous research, manipulations of these prosodic cues (duration, F0 contour, intensity) enable words manipulated at these positions to sound like boundaries[10]. Moreover, pitch and sentence-final cues must be combined to trigger the perception of intonational phrase boundaries[6]. Even auditory information representations affect syntactic ambiguity resolution[13].

The experimental materials in this study consist of spoken utterances involving "request" speech acts[19], and acoustic parameters influencing perception and anticipation of sentence ending may indicate the prosodic representation in the transition between conversational turns. This study discovers that prosodic cues at different hierarchical levels influence the perception of syntactic structures. In comparison to a single level, the predictive ability of prosodic cues at multiple levels becomes stronger. This result suggests that the processing of prosodic boundary cues is not confined solely to sentence-final words but extends to a larger prosodic level. However, this cue extraction direction is not blindly expanded but selectively focuses on the prosodic level related to the specific prosodic boundary. Although the prosodic level closer to the end of the sentence has a greater impact, prosodic cues at the earlier sentence-initial level also contribute to the listeners' perception to some extent. This could be attributed to speakers encoding from the sentence beginning and completing the encoding before starting pronunciation. It supports the idea of global encoding, further challenging the viewpoint of local encoding[18].

5. References

[1] W. Li and Y. Yang, ‘Prosodic Resolution of Syntactic Ambiguity: From the Point of View of Speaker and Listener’, *Adv. Psychol. Sci.*, vol. 15, no. 2, pp. 282–287, 2017.

[2] Y. Lin, H. Ding, and Y. Zhang, ‘Prosody Dominates Over Semantics in Emotion Word Processing: Evidence From Cross-Channel and Cross-Modal Stroop Effects’, *J. Speech Lang. Hear. Res.*, vol. 63, no. 3, pp. 896–912, Mar. 2020, doi: 10.1044/2020_JSLHR-19-00258.

[3] X. Wen, Y. Jia, and A. Li, ‘Interaction of Syntax, Semantics and Pragmatics on Discourse Prosody in Standard Chinese’, in *2018 11th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, Taipei City, Taiwan: IEEE, Nov. 2018, pp. 434–438. doi: 10.1109/ISCSLP.2018.8706640.

[4] Y. Chen, Z. Jing, G. Qian, and Y. Li, ‘Referential Status and Prosodic Encoding of the Subject in Mandarin LVS Construction’, *Foreigh Lang. Their Teach.*, vol. 01, no. 02, p. 34, 2023.

[5] Y. Zhao, ‘Explanations, Reflections and Prospects on Theories of Focus and Topic Structure at Phonology-syntax Interface’, *Foreigh Lang. Their Teach.*, no. 2, pp. 52–62, 2019.

[6] J. Holzgrefe-Lang *et al.*, ‘How pitch change and final lengthening cue boundary perception in German: converging evidence from ERPs and prosodic judgements’, *Lang. Cogn. Neurosci.*, vol. 31, no. 7, pp. 904–920, Aug. 2016, doi: 10.1080/23273798.2016.1157195.

[7] M. Wagner and D. G. Watson, ‘Experimental and theoretical advances in prosody: A review’, *Lang. Cogn. Process.*, vol. 25, no. 7–9, pp. 905–945, Sep. 2010, doi: 10.1080/01690961003589492.

[8] J. Cole, ‘Prosody in context: a review’, *Lang. Cogn. Neurosci.*, vol. 30, no. 1–2, pp. 1–31, Feb. 2015, doi: 10.1080/23273798.2014.963130.

[9] C. Männel and A. D. Friederici, ‘Neural correlates of prosodic boundary perception in German preschoolers: If pause is present, pitch can go’, *Brain Res.*, vol. 1632, pp. 27–33, Feb. 2016, doi: 10.1016/j.brainres.2015.12.009.

[10] A. Buxó-Lugo and D. G. Watson, ‘Evidence for the influence of syntax on prosodic parsing’, *J. Mem. Lang.*, vol. 90, pp. 1–13, Oct. 2016, doi: 10.1016/j.jml.2016.03.001.

[11] F. Honbolygó, Á. Török, Z. Bánréti, L. Hunyadi, and V. Csépe, ‘ERP correlates of prosody and syntax interaction in case of embedded sentences’, *J. Neurolinguistics*, vol. 37, pp. 22–33, Feb. 2016, doi: 10.1016/j.jneuroling.2015.08.001.

[12] M. M. Kjelgaard and S. R. Speer, ‘Prosodic Facilitation and Interference in the Resolution of Temporary Syntactic Closure Ambiguity’, *J. Mem. Lang.*, vol. 40, no. 2, pp. 153–194, Feb. 1999, doi: 10.1006/jmla.1998.2620.

[13] Y. Luo, M. Yan, and X. Zhou, ‘Prosodic boundaries delay the processing of upcoming lexical information during silent sentence reading.’, *J. Exp. Psychol. Learn. Mem. Cogn.*, vol. 39, no. 3, pp. 915–930, 2013, doi: 10.1037/a0029182.

[14] S. Bögels, L. Magyari, and S. C. Levinson, ‘Neural signatures of response planning occur midway through an incoming question in conversation’, *Sci. Rep.*, vol. 5, no. 1, p. 12881, Oct. 2015, doi: 10.1038/srep12881.

[15] L. Frazier, K. Carlson, and C. Cliftonjr, ‘Prosodic phrasing is central to language comprehension’, *Trends Cogn. Sci.*, vol. 10, no. 6, pp. 244–249, Jun. 2006, doi: 10.1016/j.tics.2006.04.002.

[16] I. Lehiste, ‘Phonetic Disambiguation of Syntactic Ambiguity’, *J. Acoust. Soc. Am.*, vol. 53, no. 1, pp. 380–380, Jan. 1973, doi: 10.1121/1.1982702.

[17] P. J. Price, M. Ostendorf, S. Shattuck-Hufnagel, and C. Fong, ‘The use of prosody in syntactic disambiguation’, *J. Acoust. Soc. Am.*, vol. 90, no. 6, pp. 2956–2970, Dec. 1991, doi: 10.1121/1.401770.

[18] C. Clifton, K. Carlson, and L. Frazier, ‘Informative Prosodic Boundaries’, *Lang. Speech*, vol. 45, no. 2, pp. 87–114, Jun. 2002, doi: 10.1177/00238309020450020101.

[19] J. A. Caballero, N. Vergis, X. Jiang, and M. D. Pell, ‘The sound of im/politeness’, *Speech Commun.*, vol. 102, pp. 39–53, Sep. 2018, doi: 10.1016/j.specom.2018.06.004.

[20] F. Wu, Y. Luo, and X. Zhou, ‘Building Chinese relative clause structures with lexical and syntactic cues: evidence from visual world eye-tracking and reading times’, *Lang. Cogn. Neurosci.*, vol. 29, no. 10, pp. 1205–1226, Nov. 2014, doi: 10.1080/01690965.2013.841969.

[21] J. Snedeker and J. Trueswell, ‘Using prosody to avoid ambiguity: Effects of speaker awareness and referential context’, *J. Mem. Lang.*, vol. 48, no. 1, pp. 103–130, Jan. 2003, doi: 10.1016/S0749-596X(02)00519-3.

[22] M. H. K. Ip and A. Cutler, ‘Universals of listening: Equivalent prosodic entrainment in tone and non-tone languages’, *Cognition*, vol. 202, p. 104311, Sep. 2020, doi: 10.1016/j.cognition.2020.104311.

[23] J. Krivokapić and D. Byrd, ‘Prosodic boundary strength: An articulatory and perceptual study’, *J. Phon.*, vol. 40, no. 3, pp. 430–442, May 2012, doi: 10.1016/j.wocn.2012.02.011.

[24] C. W. Wightman, S. Shattuck-Hufnagel, M. Ostendorf, and P. J. Price, ‘Segmental durations in the vicinity of prosodic phrase boundaries’, *J. Acoust. Soc. Am.*, vol. 91, no. 3, pp. 1707–1717, Mar. 1992, doi: 10.1121/1.402450.