



Building bridges: The role of prosody in Mandarin-speaking adults' and children's anaphora resolution

Kexin Du, Sergey Avrutin, Aojun Chen

Utrecht University

k.du@uu.nl, s.avrutin@uu.nl, aoju.chen@uu.nl

Abstract

Past research on the role of prosody in reference is primarily concerned with how adults and children use prosodic cues to signal accessibility change from givenness to newness of the same noun phrase. This study explores the role of prosody in the referential dependencies between one antecedent noun-phrase and one reflexive anaphor 'zi-ji' (oneself) in Mandarin-speaking adults and children. In sentences like "Boris dreamed that Miffy painted 'zi-ji'", 'zi-ji' can establish two types of anaphor-antecedent dependencies: (1) a local dependency where 'zi-ji' refers to Miffy, (2) a non-local dependency where 'zi-ji' refers to Boris. Such sentences were elicited in both interpretations from Mandarin-speaking adults and 6 to 10-year-olds in a picture-matching game. Duration analysis on 'zi-ji' shows that adults produced 'zi-ji' with a longer duration in the non-local dependency condition than in the local dependency condition. This result can be explained by the economy hierarchy model, whereby the local antecedent made more accessible by the locality constraint is preferred, thus necessitating the use of more prosodic prominence to mark the less accessible non-local antecedent. This pattern was not found in children's production, suggesting prolonged acquisition of using prosody to build anaphor-antecedent dependencies for 'zi-ji'.

Index Terms: prosody, reflexive pronoun zi-ji, Mandarin, syntax-discourse interface, acquisition

1. Introduction

1.1. Background

Building a fluent discourse requires incrementally and repeatedly building references to the same referential entities. Among a variety of referential forms available, such as noun-phrases (like "the rabbit"), pronouns (like "her"), reflexives (like "herself"), the reference to the same referential entity ("Miffy") in discourse can be established via anaphor-antecedent dependencies between these expressions without having to repeatedly use the proper name. As the discourse progresses, more referential entities are introduced (as antecedent candidates), the activation (accessibility) level of various individual referential entities (hereafter referents) thus fluctuates. Much effort has been devoted to characterizing how this affects the lexical and prosodic encoding of the anaphoric expressions. The general tendency seem to be one of reverse relation, the less accessible the referent is, the more prominent (unreduced) the anaphoric expression is [1-3].

In speech production, to establish anaphor-antecedent dependencies using the appropriate prosodic form in accordance with the referent's accessibility level requires

sophisticated coordination of various linguistic modules. Specifically, interlocutors are expected to know (a) all the syntactic and extra-syntactic constraints of different anaphors to figure out and further keep track of the accessibility levels of all the introduced antecedent candidates, (b) whether prosodic prominence is needed in case the intended antecedent is relatively less accessible than others, and eventually (c) which prosodic dimensions among pitch, intensity and duration are available for manipulation. Although children younger than 2 years are already sensitive to both prosodic variation and referential accessibility [4], it still takes years for them to master the mapping of the two.

Previous studies on children's production of prosody in reference have shown that children of different languages develop in stages the ability to prosodically attune the referential expression to accessibility status shifts from given/accessible to new/inaccessible [5]. For example, English-speaking 3- and 4-year-olds can produce a larger pitch span to distinguish referents that differ in both contrastiveness and referential givenness-newness, but cannot use durational cues like adults [6-8]. Dutch-speaking children can vary type of pitch accent to distinguish referents differing in givenness-newness in adult-like way at the age of 7 or 8. They can also vary the phonetic realization of a pitch accent in terms of pitch range for this purpose at the age of 7 or 8 but varying the use of duration is not acquired at the age of 7 or 8 [9-10]. In Mandarin, a language with lexical tones, whereas adult use both pitch and duration to encode referential givenness-newness [11], children can use durational cues in all four lexical tones and pitch cues in some lexical tones in an adult-like way at the age of 4 or 5, and the use of pitch cues is not fully adult-like at the age of 11 [12].

These studies were almost exclusively focused on the same noun-phrase experiencing referential accessibility changes in givenness-newness, not involving how the accessibility level of one antecedent noun-phrase alters the prosodic prominence level of the subsequent anaphoric expression like pronouns and reflexives.

1.2. The current study

A strict reflexive is different from pronouns in that its dependency can be established within the syntactic module. According to Chomsky's Canonical Binding Principles [13], while a pronoun finds its antecedent outside the local (minimal) syntactic domain containing the pronoun like in (1), a reflexive is bound to the antecedent within the local syntactic domain containing the reflexive (e.g., the same clause) like in (2).

(1) Miffy's mom₁ dreamed that Miffy₂ painted her_{1/*2}

(2) Miffy's mom₁ dreamed that Miffy₂ painted herself_{*1/2}

(3) Boris₁ dreamed that Miffy₂ painted zi-ji_{1/2}

Notably, the Mandarin reflexive ‘zi-ji’ (roughly standing as “oneself”) behaves differently from strict reflexive elements. As shown in (3), rather than being bound only to the local antecedent “Miffy”, ‘zi-ji’ can also refer to the antecedent “Boris” outside of its local syntactic domain (non-local). Two types of referential dependencies can thus be established between ‘zi-ji’ and its antecedent: (a) a local reflexive-like dependency between ‘zi-ji’ and “Miffy”, (b) a non-local pronoun-like dependency between ‘zi-ji’ and “Boris”.

It has been recently proposed that this ambiguity is due to the bi-morphemic nature of ‘zi-ji’ which contains a reflexivizer prefix ‘zi-’ and a pronominal stem ‘-ji’. The local dependency results from ‘zi-’ reflexivizing the local verb, whereas the non-local dependency results from ‘-ji’ taking over the argument, probing outside the local domain like a pronoun [14]. And since ‘-ji’ has no gender feature, ‘zi-ji’ can disregard gender restrictions in choosing the antecedent.

Crucially, according to the economy hierarchy model [15], between the two antecedent candidates, the local one is claimed to be more accessible over the non-local one due to the fact that the reflexive ‘zi-ji’ and the local antecedent share the immediate verb grid (locality). For ‘zi-ji’ to probe further outside its own syntactic domain would require more processing resources, making the non-local antecedent less accessible [16–17]. A recent study on the acquisition of ‘zi-ji’ in Mandarin-speaking children shows that 3 to 6-year-olds can already interpret these two dependencies and they excel at building the non-local dependency, showing adult-like judgments as early as 4 years old [18].

Two questions then arise: (i) Do Mandarin-speaking adults distinguish, by means of prosodic prominence, the accessibility difference caused by such syntactic constraint of locality between these two antecedent candidates of ‘zi-ji’? (ii) Are Mandarin-speaking children similar to adults in this respect? The prosodic cue to be examined in this study is duration. Based on the claim in [1–3], we hypothesize that adults will produce ‘zi-ji’ with a longer duration when referring to a non-local antecedent than when referring to a local antecedent in Mandarin. In the light of past work on children’s use of durational cues in reference (albeit only in NPs) and children’s knowledge of legitimate antecedents of ‘zi-ji’, we hypothesize that children aged 6 and older can use duration to distinguish the two types of anaphora-antecedent dependencies, like adults. These hypotheses were tested in a pilot study on Mandarin-speaking adults and children.

2. Methodology

2.1. Participants

Five Mandarin-speaking adults (2 females and 3 males, ages 24–30, mean age 27) and five Mandarin-speaking children (2 females and 3 males, ages 6–10, mean age 8) participated in the experiment. The adult participants were students at Utrecht University at the time of testing. Brought up in China, they were native Mandarin speakers with high proficiency. Due to travel restrictions, children participants were recruited in the Netherlands. They were balanced Mandarin-Dutch bilinguals. They spoke fluent and accent-less Mandarin Chinese with family members in their households. All participants were tested in a quiet room at their homes after school time. They received either a gift or financial compensation.

2.2. Materials

We adopted [12]’s picture matching game to elicit utterances with ‘zi-ji’. In this game, the participant was supposed to look at pictures and describe the event depicted in one of the pictures to the experimenter using a fixed target sentence structure as illustrated above (3).

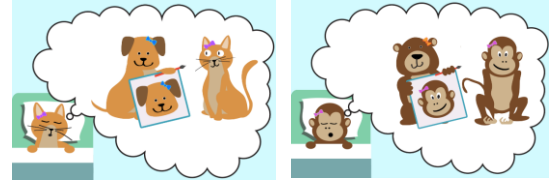


Figure 1: *Picture event in local (left panel) and non-local conditions (right panel).*

As illustrated in Figure 1, the picture events depicted animal A dreaming about animal B doing certain actions to either (a) oneself (‘zi-ji’ as animal B, local condition); or (b) to the dreamer (‘zi-ji’ as animal A, non-local condition).

10 relativizable verbs that were familiar to Mandarin 3-year-olds in both production and perception were selected from Wordbank [19]. In total, 20 target picture stimuli are created. Another 30 unrelated fillers are used to guise the purpose of study. Each one verb occurred in two locality conditions with two different pairs of animals: ‘zi-ji’ had the local interpretation in one verb + animal pair combination and the non-local interpretation in another verb + animal pair combination (see Figure 1 for verb “paint”). The occurrence of animal figures was balanced in terms of the subject/object positions. Each animal pair was manipulated to have similar visual prominence (e.g. equal size, same color scheme).

2.3. Procedure

In the adapted picture-matching game, 3 piles of pictures were used. The experimenter held one pile of pictures with missing parts; another pile of pictures containing the missed part were laid around on the table seemingly ‘messy’; and the participant had one pile of pictures containing all the information on the computer screen not visible to the experimenter. The participant’s task was to help the experimenter put the two piles of incomplete pictures into matched pairs by describing the complete pictures on the computer screen in response to the experimenter’s question.

On every trial, the experimenter showed a picture from her pile to the participant, described the picture and asked a question about the missing part, as illustrated below. Following the participant’s description, the experimenter then went on to find the matching picture in the ‘messy’ pile.

Exp: 看! 小猴梦见小熊, 小猴梦见小熊怎么了?

Look! Little monkey dreamed of little bear. Little monkey dreamed little bear what? (What did the little monkey dream that the little bear did?)

Child: 小猴梦见小熊画了自己。

Little monkey dreamed little bear painted ‘zi-ji’.

To ensure the choice of words in participants’ production, we introduced a virtual robot to the game as their informant. The robot provided participants with the ‘raw’ answers to each question using the intended sentence structure via a headphone set. Crucially, sentence-level prosodic features were not

present in the robots' sentences. The words in the robots' answers were first recorded in a randomized wordlist by a female native speaker of Mandarin Chinese. The words' original pitch was erased and was re-set at 400 Hz using Praat [22]. These words were then spliced together with a 100 ms pause in between to form the Mandarin target sentences.

The participants were instructed to speak normally but use the exact wording of the robot. Each session lasted about 30-50 minutes with children participants and about 30 minutes with adult participants. The sessions were recorded by means of a portable ZOOM H1 Handy recorder at 48kHz sampling rate with 16 bit resolution.

Prior to the picture matching game, to familiarize participants with the robotic sounding of these terms, the images of all the animal figures and the actions were first introduced by the virtual robot. A short picture-naming task was then conducted to make sure that the participants would use intended words to refer to the animals and the actions.

2.4. Data annotation

None of the participants used the robotic voice, and few gave elided answers. One child's recording was excluded due to heavy constant background noise. Usable sentences were selected from the remaining recordings. Deviant responses were excluded from further analysis. A target sentence was considered deviant in the following cases: (1) incomplete sentence (word insertion, deletion, or replacement); (2) repetition or self-repair; (3) perceivable hesitation; (4) too much background noise; (5) answers before the experimenter asks the question. In total, 99 sentences from the 5 adults and 65 sentences from 4 children were included in the analysis. Each selected sentence was annotated for the boundaries of 'zi-ji' in Praat.

3. Results

Mixed-effects modelling in the program R with the lme4 package [23] was used to assess 2 sets of data (adults and children). The duration of 'zi-ji', the lengthening proportion of 'zi-' in 'zi-ji', and the lengthening proportion of '-ji' in 'zi-ji' were analyzed as dependent variables. The latter two were included to explore whether the two morphemes of 'zi-ji' might be lengthened to different degrees in different conditions.

The random factors were SPEAKER (i.e. the participants) and TRIAL (i.e. the items). The fixed factor was LOCALITY (local vs. non-local). Two models were built, one with only the random factor, and one with both the random factor and the fixed factor. The ANOVA function in R was used to compare these models. A statistically significant difference between these two models would indicate a main effect of the fixed factor.

3.1. Duration of 'zi-ji'

Our analysis revealed that (1) There was a main effect of LOCALITY in the Mandarin-speaking adults ($\chi^2=4.9394$, $p=0.02625$), with non-local dependency increasing the duration of 'zi-ji' by about $33.8 \text{ ms} \pm 14.82 \text{ (SE)}$; (2) There was no main effect of LOCALITY in the data of Mandarin-speaking children ($p=0.9442$).

However, close inspection of the children's data showed that while most children did not prosodically distinguish the two locality conditions, there was one child (participant 2, age

8 years old) who could use lengthening 'zi-ji' in the non-local condition, albeit of slower manner than adults (see Figure 3).

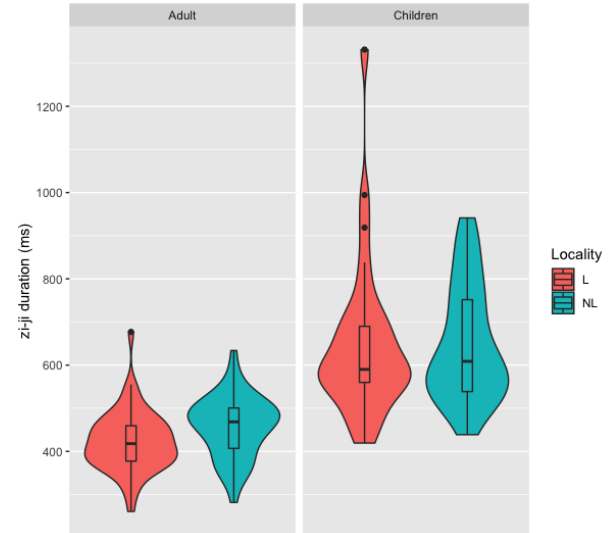


Figure 2: Comparison of 'zi-ji' duration under two conditions between adults and children.

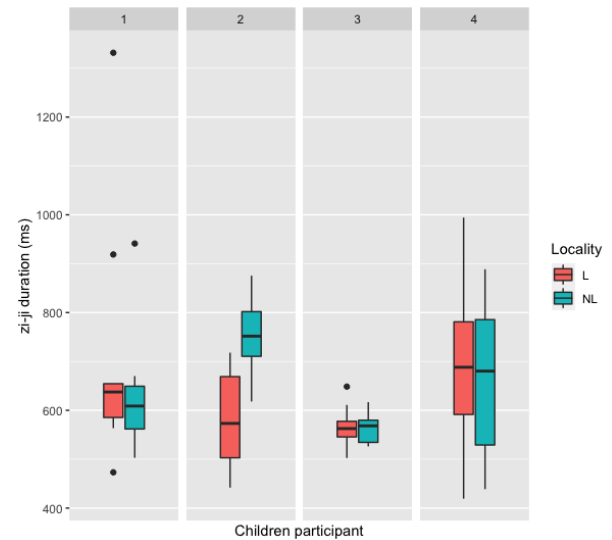


Figure 3: Comparison of 'zi-ji' duration among children participants

3.2. Lengthening proportions of 'zi-' and '-ji' in 'zi-ji'

Due to the relatively small sample size, no inferential result can be drawn from the analysis regarding the ratio of the two elements. We thus present the descriptive statistics.

As illustrated in Figure 4, the lengthening proportions of 'zi-' in 'zi-ji' are very similar in the local dependency condition and the non-local dependency for both adults and children (adult's mean for L: 43%, and for NL: 40%; children's mean for L: 37%, and for NL: 35%).

Similarly, the proportion of '-ji' in 'zi-ji' in the two conditions for both adults and children are very much alike, as can be seen in Figure 5 (adult's mean for L: 56%, and for NL: 59%; children's mean for L: 61%, and for NL: 63%).

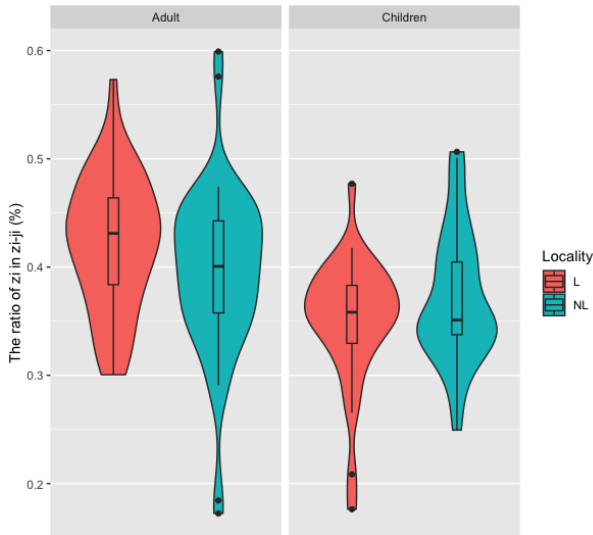


Figure 4: Comparison of the lengthening proportion of 'zi-' between groups

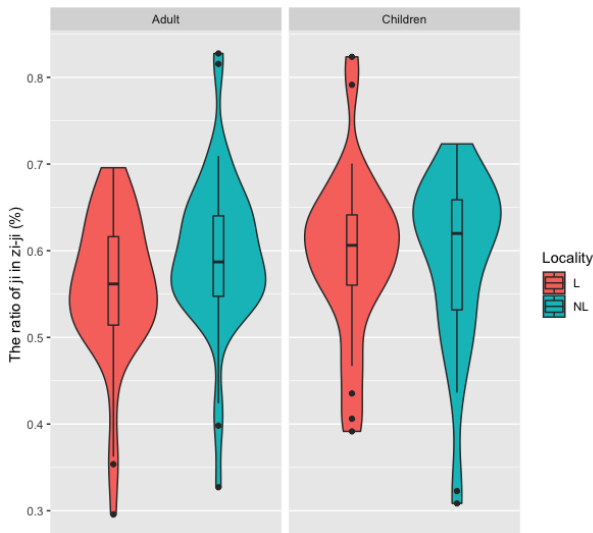


Figure 5: Comparison of the lengthening proportion of '-ji' between groups.

4. Discussion and conclusions

This study aims at investigating whether Mandarin-speaking adults distinguish the accessibility difference caused by the syntactic constraint of locality between the two antecedent candidates of 'zi-ji' via prosodic prominence, and whether Mandarin-speaking children show adult-like performance.

In line with our hypothesis, we have found that Mandarin-speaking adults did produce 'zi-ji' with a longer duration when referring to a non-local antecedent than when referring to a local antecedent. These results suggest there is indeed an interplay between the referential accessibility difference (brought by the syntactic constraint of locality) and the prosodic realization of the Mandarin reflexive anaphor 'zi-ji'. By highlighting 'zi-ji' with more prosodic prominence (via lengthening the whole anaphor) the anaphor is licensed to build its dependency with the less accessible non-local

antecedent. We did not find a significant difference between the lengthening distribution over the two morphemes of 'zi-ji' in non-local condition.

In the light of past work on children's use of durational cues in reference and children's knowledge of legitimate antecedents of 'zi-ji', we also hypothesized that children aged 6 and older could use duration to distinguish the two types of anaphor-antecedent dependencies, like adults. However, we found no evidence that they vary either the duration of 'zi-ji' or the lengthening of 'zi-' or '-ji' based on the referential accessibility of the two antecedent candidates brought by the syntactic constraint of locality. Although among the four children, one child (8 years old) did show adult-like pattern.

These results suggest that children do not seem to distinguish the two antecedents in terms of referential accessibility brought by locality by the age of 10. This appears to be in line with [22-23] which hypothesized that at some stage of child's development, the syntactic and extra-syntactic routes of building dependencies may not differ in terms of required processing resources. Since studies have shown that 4 year-olds Mandarin children already show adult-like behavior in building non-local dependency (which is not favored by the locality constraint of the syntax module), it is reasonable to speculate that the referential candidate made more accessible by extra-syntactic modules might be the preferred antecedent by children. Furthermore, our results may also imply that children have not fully acquired the use of prosody to encode anaphora-antecedent dependencies in Mandarin Chinese by the age of 10. Since Mandarin-speaking children can already correctly interpret these two types of dependencies for 'zi-ji' at the age of four, their failure to use prosody to realize the different types of dependencies may be attributed to limited ability of manipulating the durational property of 'zi-ji' in different conditions as a reflexive element, in spite that they can do this in the case of noun-phrase referents. This limitation might be related to that anaphoric expressions in general are typically not pronounced with high level of prosodic prominence.

However, the results of this study need to be validated by a larger set of data from adults and from monolingual Mandarin-Chinese speaking children. They serve as first indicators for a larger-scale study on the role of prosody (i.e. duration, pitch, intensity) in building referential bridges between 'zi-ji' and its antecedent in both child and adult Mandarin Chinese.

5. Acknowledgements

We would like to thank all the participants for their willingness to help during the pandemic, Jianan Liu for the audio stimuli recording, Kirsten Schutter for the statistical support, and Ingo Caly for the cartoon illustrations.

6. References

- [1] M. Ariel, *Accessing Noun-Phrase Antecedents*. Croom Helm Linguistics Series. London: Routledge, 1990.
- [2] J. E. Arnold, "How speakers refer: The role of accessibility," *Language and Linguistics Compass* vol. 4, no. 4, pp. 187-203, 2010
- [3] D. G. Watson, "The many roads to prominence: Understanding emphasis in conversation." *Psychology of Learning and Motivation*, vol. 52, pp. 163-183, 2010.
- [4] N. Esteve-Gibert and P. Prieto. "Early development of prosody-meaning interface," in *The Development of Prosody in First*

Language Acquisition, P. Prieto and N. Esteve-Gibert, Eds. 2018. pp. 228-246.

- [5] A. Chen, "Children's use of intonation in reference and the role of input," in *The Acquisition of Reference*, L. Serratrice, and S.E. Allen, Eds. John Benjamins Publishing Company, 2015, pp. 83-104.
- [6] P. A. Homby and W. A. Hass. "Use of contrastive stress by preschool children." *Journal of Speech and Hearing Research*, vol. 13, no. 2, pp. 395-399. 1970.
- [7] B. MacWhinney and E. Bates. "Sentential devices for conveying givenness and newness: A cross-cultural developmental study." *Journal of Verbal Learning and Verbal Behavior*, vol. 17, no. 5 pp. 539-558. 1978.
- [8] E. Wonnacott and D. G. Watson. "Acoustic emphasis in four year olds." *Cognition*, vol. 107, no. 3, pp. 1093-1101, 2008.
- [9] A. Chen, "The phonetics of sentence-initial topic and focus in adult and child Dutch," in *Phonetics and Phonology: Interactions and Interrelations*, M. Vigário, S. Frota and M. J. Freitas, Eds. Amsterdam: Benjamins, 2009. pp. 91-106.
- [10] A. Chen, "Tuning information packaging: intonational realization of topic and focus in child Dutch". *Journal of Child Language*, vol. 38, pp. 1055-1083, 2011.
- [11] I. C. Ouyang and E. Kaiser. "Prosody and information structure in a tone language: an investigation of Mandarin Chinese." *Language, Cognition and Neuroscience*, vol. 30, no. 1-2, pp. 57-72, 2015.
- [12] A. Yang and A. Chen. "The developmental path to adult-like prosodic focus-marking in Mandarin Chinese-speaking children." *First language*, vol. 38, no. 1, pp. 26-46. 2018.
- [13] N. Chomsky. *Lectures on government and binding*. Dordrecht: Foris. 1981.
- [14] E. Reuland, S. H. Wong, and M. Everaert. "How the complexity of Mandarin zi-ji simplifies the grammar." *Linguistic Inquiry*, vol. 51, no. 4, pp. 799-814. 2020.
- [15] T. Reinhart, *Anaphora and Semantic Interpretation*. London: Croom Hem, 1983.
- [16] S. Zhang, D. Zhang, J. Wu, L. Zhao and T. Guo, "Who is 'oneself' in Chinese? ERP responses to the Chinese simple reflexive ziji in discourse comprehension." *Journal of Neurolinguistics*, vol. 58, no. 100961, 2021.
- [17] X. Li and X. Zhou, "Who is ziji? ERP responses to the Chinese reflexive pronoun during sentence comprehension." *Brain Research*, no. 1331, pp. 96-104, 2010.
- [18] Y. Xie, S. Avrutin and P. Coopmans. "Development of Syntax-Discourse Interface Knowledge in 3-to 6-Year-Old Mandarin Chinese Speaking Children." *Proceedings of the 44th Boston University Conference on Language Development*, pp. 738-748. 2020.
- [19] M. C. Frank, M. Braginsky, D. Yurovsky and V. A. Marchman. "Wordbank: An open repository for developmental vocabulary data." *Journal of Child Language*, vol. 44, no. 3, pp. 677-694. 2017.
- [20] P. Boersma, "Praat, a system for doing phonetics by computer," *Glott International*, vol. 5, no. 9, pp. 341-345, 2001.
- [21] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4", *J. Stat. Soft.*, vol. 67, no. 1, pp. 1-48, Oct. 2015.
- [22] S. Avrutin, *Development of the Syntax-Discourse Interface*. Dordrecht: Kluwer Academic, 1999.
- [23] S. Avrutin, "Weak Syntax," in *Broca's Region*, Y. Grodzinsky, and K. Amunts, Eds. New York: Oxford University Press, 2006, pp. 49-62.