

Children’s Acquisition of Mandarin Tone 3 Sandhi in Flat Structures

Chiung-Yao Wang, Cristina Schmitt, Yen-Hwei Lin

Department of Linguistics and Languages, Michigan State University, USA

wangch10@msu.edu, schmit12@msu.edu, liny@msu.edu

Abstract

The goal of this study is to examine children’s acquisition of Mandarin Tone 3 Sandhi (T3S) in flat structures. Syllables in flat structures are parsed either in binary feet from left to right, followed by incorporation of an unparsed syllable or in one large prosodic domain [1-3]. We used an elicited production task to examine non-cyclic parsing strategies in T3S application in sequences of two, three, and five digits. 46 children (ages 3 & 5) and 20 adults were tested. The results show that children could apply T3S non-cyclically in digit sequences, although there were over- and under-application errors. In a five-digit sequence, adults (but not children) produced the predicted binary foot followed by a ternary foot. All age groups used an unpredicted pattern made of a ternary foot followed by a binary foot, which challenges the previous account [1-3]. We argue that there is a clear developmental pattern in children’s acquisition of T3S in flat structures, and suggest that foot-building in digits does not always follow the typical foot-building process.

Index Terms: Mandarin Tone 3 Sandhi, flat structures, prosodic domains, non-cyclic parsing, child language acquisition

1. Introduction

T3S changes a Tone 3 (T3, low dipping tone) to a Tone 2 (T2, mid rising tone) when it is followed by another Tone 3 [1-3]. T3S requires setting up the prosodic domains within which T3S applies so that adjacency can be defined. Typically, syntax and prosody both play vital roles in the building of such domains. In a flat structure, there is no compositional building of a tree, so all the syllables are arguably not hierarchically organized. A flat structure therefore serves as a perfect opportunity to investigate T3S with the focus on the prosody alone.

To our knowledge, there are no previous studies on children’s acquisition of T3S in flat structures (e.g. polysyllabic names and sequences of digits) but previous studies suggest early acquisition of Mandarin Tone 3 Sandhi (henceforth T3S) [4-8].

Our study examines how children build prosodic domains, in the context of two, three, and five digits. What does T3S depend on when there is “no syntax” for it to refer to? The issues we are concerned with in T3S application in flat structures are the following: (i) foot building; (ii) incorporation of syllables that have not been parsed into foot structures; (iii) directionality of foot building. In addition, we are interested in additional parsing patterns in adults and children and also in children’s developmental pattern in the acquisition of T3S in flat structures and what it tells us about children’s way of grouping syllables into larger units.

1.1. T3S in flat structures

Syllables in flat structures are parsed either in disyllabic feet from left to right [1-3, 9, 10], followed by incorporation of an unparsed syllable into its preceding foot, or in one large

prosodic domain [1-3]. A disyllabic foot is a preferred foot structure in Mandarin Chinese, and such binary foot comprises the basic T3S domain within which T3S must apply [11]. For the binary pattern, three elements are crucial: (i) binary foot-building, (ii) left to right parsing, and (iii) incorporation of unparsed syllables. For convenience, we refer to these three crucial elements as *L-to-R-Bin-Incorp*. The example in (1) shows a five-digit string, which is parsed into two disyllabic feet, followed by incorporation of the unparsed syllable (ST1). The larger domain pattern (ST2), which applies T3S iteratively from left to right within one single prosodic domain, is regarded as an alternative secondary fast speech pattern [1-3, 11] or variant pattern [9, 10].

(1)	wu	wu	wu	wu	wu	
	five	five	five	five	five	‘five-five-five-five-five’
	T3	T3	T3	T3	T3	UT(= underlying tones)
	(T2	T3)	T3	T3	T3	
	(T2	T3)	(T2	T3)	T3	
	(T2	T3)	(T2	T2	T3)	ST1 (= surface tones 1)
	(T2	T2	T2	T2	T3)	ST2 (= surface tones 2)

(A boldtype T2 indicates a derived sandhi tone.)

Results from an acoustic experimental study [12] where adult participants were asked to produce sentences at three speech rates, slow, normal, and fast show that both *L-to-R-Bin-Incorp* and larger domain parsing strategies are attested in slow and fast speech rates.

1.2. Research questions, hypotheses, and predictions

Our study examines how children and adults parse strings of digits prosodically and whether or not they produce the predicted patterns in (1). We ask the following questions: (i) When the larger domain pattern is not used, is binary parsing the main foot-building strategy? (ii) In the *L-to-R-Bin-Incorp* pattern, is an unfooted syllable incorporated into a neighboring foot in odd number of syllables? (iii) What is the directionality of foot-building in flat structures? (iv) Is *L-to-R-Bin-Incorp* the dominant parsing strategy in flat structures?; (v) Is there a developmental pattern in children’s acquisition of T3S in flat structures?

We put forth four hypotheses regarding questions (i-iv) (H1 – H4) [1-3, 9, 10]: **Binary parsing (H1):** Binary feet are built iteratively until no more binary foot can be built; **Incorporation (H2):** If there is an unparsed syllable, it is incorporated into a neighboring foot; **Directionality L to R (H3):** Binary feet are formed from left to right (based on T3S models [1-3, 9-10]); **Binary dominance (H4):** *L-to-R-Bin-Incorp* is the dominant parsing strategy. In addition, children may have structurally different domains in production planning due to their more limited processing resources [13]. We hypothesize that children may not produce larger domain parsing as much as adults can, and that their prosodic strategies and T3S acquisition develop with time. **Prosodic and T3S development (H5)** is put forth: Children produce

less larger domain parsing than adults and T3S develops with age.

Predictions for sequences of two, three, and five T3-digits are (i) (T2T3), (ii) (T2T2T3), and (iii) (T2T3)(T2T2T3) or (T2T2T2T2T3), and *L-to-R-Bin-Incorp* is the dominant parsing strategy. In addition, we expect children to produce the larger domain pattern less frequently than adults and 5-year-olds have a higher correct rate than 3-year-olds.

2. Methodology

2.1. Participants

Sixty-six subjects were recruited in Taichung, Taiwan for this study. Table 1 shows the distribution of three age groups: 3-year-olds, 5-year-olds, and adults (as the control group).

Table 1. Distribution of the subjects

Age groups	N	Age range	Mean	s.d.
3-year-olds	19	3;4 – 3;11	4;4	2.42 (mo.)
5-year-olds	27	5;1 – 5;11	6;3	3.05 (mo.)
adults	20			

2.2. Materials

Digits from 0 to 9 are all single syllables in Mandarin. Except for “5” and “9” which are in Tone 3, all the rest of the digits are in Tone 1, Tone 2, or Tone 4 (i.e. non-T3s). T3-digits “5” and “9” were used as the test items, and non-T3 digits were used as the control items and in the practice session.

In the control items, the surface tones and the underlying tones are the same because non-T3s are not affected by T3S. In the test items, T3S applies according to how the string of syllables is parsed. Surface tones differ from underlying tones due to T3S application. A non-T3 digit *si* (T4) ‘four’ was used in the practice session. Two T3-digits *wu* ‘five’ and *jiu* ‘nine’ were used in two-, three-, and five-digit sequences as the test items (i.e. 55, 555, 55555; 99, 999, 99999). Two non-T3 digits *er* (Tone 4) ‘two’ and *san* (Tone 1) ‘three’ were used in two-, three-, and five-digit sequences as the control items (i.e. 22, 222, 22222; 33, 333, 33333).

2.3. Procedure

An elicited repetition task [14, 15] was used. All children and adults were tested individually in a quiet room. As the digit appeared on the screen of a laptop computer, the experimenter asked the child what it was. This was to make sure that the child knew the digit and could say it with the underlying tone correctly. The experimenter said, “What’s this?” (pointing to the digit on the screen). After the child gave the answer, she was told to hold out one hand just like the experimenter showed her, with five fingers up straight. Then the experimenter gently bent down three of her fingers, leaving two up and said, “You say it (pointing to the digit on the screen) when I tap your fingers, okay?” As two fingers were up, the child said the digit upon each of the two fingers was tapped by the experimenter’s index finger. The same procedure was followed for the 3- and 5-digit sequences, although the numbers of fingers shown differ.

Adults also saw the digit on the computer screen, but were instructed to say the digit two, three, and five times. All subjects’ responses were recorded on a Marantz PMD660 with an Audio-technica miniature clip-on microphone (AT831B Cardioid Condenser Lavalier microphone).

3. Results

3.1. Controls and test items

We first present the results of the control items in Table 2. It is clear that saying the non-T3 digit for five times in the experiment poses no problem for children.

Table 2. Control items (non-T3 digits)

Number of syllables	σσ % (N)	σσσ % (N)	σσσσσ % (N)
3-year-olds	100 (33/33)	100 (33/33)	100 (37/37)
5-year-olds	100 (54/54)	100 (54/54)	100 (54/54)
Adults	100 (40/40)	100 (40/40)	100 (40/40)

For T3-digits, while adults did well in the test items (97.50% correct in two, three, and five syllables), children’s correct rates dropped dramatically for T3 sequences (Fig. 1).

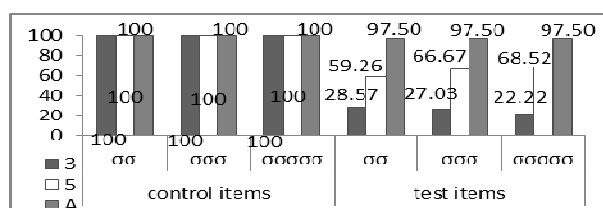


Figure 1: Correct rates in control and test items by age groups

Since children did perfectly in the control items, T3S is the source of difficulties which caused the dropping of correct rates in both child groups. There is only one acceptable surface pattern, (T2T3) and (T2T2T3) for two-syllable and three-syllable test items, respectively. These predicted patterns match what adults produced, and were also attested in children. For five-syllable items, the predicted patterns are (T2T3)(T2T2T3) and (T2T2T2T2T3), with the former pattern attested in adults only, and the latter pattern attested in all age groups. An additional pattern (T2T2T3)(T2T3) was found in adults as well as children. (See Table 3.)

Table 3. Correct rates (%) in test items (T3 digits)

T3-digits	σσ	σσσ	σσσσσ			Total
	(23)	(223)	(22223)	(23)(223)	(223)(23)	
3	28.57 (10/35)	27.03 (10/37)	16.67 (6/36)	0 (0/36)	5.56 (2/36)	22.22
5	59.26 (32/54)	66.67 (36/54)	61.11 (33/54)	0 (0/54)	7.41 (4/54)	68.52
A	97.50 (38/40)	97.50 (39/40)	70 (28/40)	22.50 (9/40)	5 (2/40)	97.50

The most common surface pattern in all age groups is the larger domain pattern. Five-year-olds’ correct rates are below 70% and they are far from adult-like. The correct rate of about 20% shows that 3-year-olds had a lot of difficulties with T3S in the test items. The predicted (T2T3)(T2T2T3) is missing in both child groups.

Logistic regression analyses were conducted for correct and incorrect responses in flat structures. The results for two-, three-, and five-digits will be presented separately.

Two T3-digits: The results show that age is significant (chi square = 46.067, $p < .001$ with $df = 2$). For correct surface pattern T2T3 relative to errors, both 3-year-olds and 5-year-olds are significantly different from adults in T3S application

in two T3-digits and they are less likely than adults to have the correct surface pattern of **T2T3** (3-year-olds: Odds Ratio (OR) = .010, $p < .001$; 5-year-olds: OR = .037, $p = .002$). There is also a significant difference between 3-year-olds and 5-year-olds (OR = .275, $p = .006$).

Three T3-digits: The results show that age is significant (chi square = 48.539, $p < .001$ with $df = 2$). For correct surface pattern **T2T2T3** relative to errors, both 3- and 5-year-olds are significantly different from adults in T3S application in three T3-digits (3-year-olds: OR = .009, $p < .001$; 5-year-olds: OR = .051, $p = .005$). There is a significant difference between 3-year-olds and 5-year-olds (OR = .185, $p < .001$).

Five T3-digits: The results show that age is significant (chi square = 71.132, $p < .001$ with $df = 6$). For five T3-digits, three surface patterns were attested in adults—larger domain parsing (**22223**), Binary-Ternary parsing (**23**)(**223**), and Ternary-Binary parsing (**223**)(**23**). The last pattern, Ternary-Binary parsing, is not predicted by *L-to-R-Bin-Incorp*, but was attested in all age groups with a low frequency (3-year-olds: 5.56%, 5-year-olds: 7.41%, and adults: 5%). For larger domain parsing (**22223**) relative to errors, 3-year-olds and 5-year-olds are found to be significantly different from adults, and both child groups are less likely than adults to use the larger domain parsing strategy (3-year-olds: OR = .008, $p < .001$; 5-year-olds: OR = .069, $p = .012$). Three-year-olds and 5-year-olds are significantly different (OR = .110, $p < .001$). For Ternary-Binary parsing—(**223**)(**23**) relative to errors, 3-year-olds are found to be significantly different from adults (OR = .036, $p = .020$) while 5-year-olds are not (OR = .118, $p = .112$). The two child groups are not significantly different from each other (OR = .304, $p = .195$).

3.2. Children's T3S Errors

The T3S error analysis is focused on children's errors by comparing 3-year-olds' errors to 5-year-olds'.

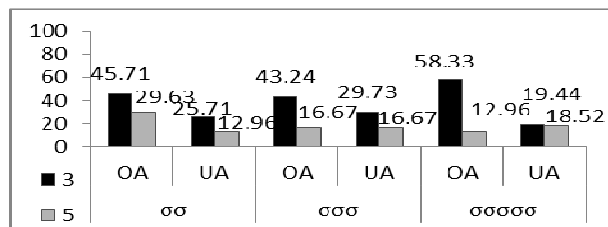


Figure 2: Children's error rates by type in flat structures (OA: Over-application; UA: Under-application)

In Figure 2, there is a clear developmental trend. The error rates decrease by age, regardless of the error types. Three-year-olds are prone to make over-application errors. Five-year-olds' T3S errors do not show a strong tendency of over- or under-application in three- and five-syllable items. In two-syllable items, however, they tend to over-apply T3S.

Logistic regression analyses were conducted for children's error types in flat structures.

Two T3-digits: Age is significant (chi square = 7.447, $p = .024$ with $df = 2$). For both error types relative to correct surface pattern (**T2T3**), 3-year-olds are significantly different from 5-year-olds (Over-application: OR = 3.100, $p = .026$; Under-application: OR = 3.986, $p = .026$). The Odds Ratio value indicates that 3-year-olds are about three times more likely than 5-year-olds to over-apply T3S rule. They are four times more likely than 5-year-olds to under-apply T3S.

Three T3-digits: Age is significant in three T3-digits (chi square = 14.592, $p = .001$ with $df = 2$). For both error types relative to correct surface pattern (**T2T2T3**), 3-year-olds are significantly different from 5-year-olds (Over-application: OR = 6.400, $p = .001$; Under-application: OR = 4.400, $p = .010$). The Odds Ratio value indicates that 3-year-olds are 6.5 times more likely than 5-year-olds to over-apply T3S, and they are 4.5 times more likely than 5-year-olds to under-apply T3S.

Five T3-digits: The results show that age is significant (chi square = 24.496, $p < .001$ with $df = 2$). For Over-application, 3-year-olds are significantly different from 5-year-olds (OR = 13.875, $p < .001$). The Odds Ratio value indicates that 3-year-olds are roughly 14 times more likely than 5-year-olds to over-apply T3S. The two child groups are not significantly different in the error type, Under-application (OR = 3.237, $p = .062$).

4. Discussion

The adult T3S patterns attested in this study are compared against the surface patterns predicted by *L-to-R-Bin-Incorp*. **Binary parsing (H1)** and **Incorporation (H2)** are supported by adults' answers of (**T2T3**) and (**T2T2T3**) in the two and three T3-digit items respectively, just as predicted. No adults produced two T3s in the two-syllable items. The fact that (**T2T3**) was the only response in the adult group indicates that a binary foot is formed for two syllables. For three-syllable items, if we had the answer type (**T2T3**)(**T3**) or (**T2T3**)(**T2**), it would be evidence against **Incorporation (H2)**, but these patterns were not attested in adults.

For testing **H3, Directionality L to R**, five T3-digits were used. We cannot use three T3-digits to test directionality, partially because we were unable to disambiguate the sources of (**T2T2T3**). This pattern can come either from (i) the *L-to-R-Bin-Incorp* parsing or (ii) the larger domain parsing; however, the unattested pattern (**T3T2T3**) in adults sheds some light on this issue because (**T3T2T3**) is a pattern that results from right-to-left parsing, followed by incorporation of the first syllable.

In five T3-digits, the larger domain parsing (**T2T2T2T2T3**) is the dominant pattern across age groups (adults: 70%, 5-year-olds: 61.11%, and 3-year-olds: 16.67%). Adults prefer (**T2T2T2T2T3**) to (**T2T3**)(**T2T2T3**), and children show a very strong preference of (**T2T2T2T2T3**) and produced no (**T2T3**)(**T2T2T3**). A possible interpretation of the result is that treating the sequence of five identical digits as an unanalyzed chunk (i.e. grouping them in one large domain) might be less marked in processing than segmenting the sequence through the process of foot-building (*L-to-R-Bin-Incorp*, resulting in a binary foot followed by a ternary foot). The former involves **T2** (Mid-High; MH) to be produced four times and end with **T3** (Low; L), and the latter involves producing the sequence of more alternations between the two tones (MH-L-MH-MH-L), and hence may be more marked in terms of production.

The predicted pattern (**T2T3**)(**T2T2T3**) was attested only in the adult group, at 22.50%. Not a single child produced this pattern. The fact that (**T2T3**)(**T2T2T3**) was attested, but not (**T3T2T3**)(**T2T3**), gives strong evidence that left-to-right parsing, rather than right-to-left parsing, was used. **Directionality L to R (H3)** is supported by the adult data. In addition, (**T2T3**)(**T2T2T3**) confirms **Incorporation (H2)** that an unfooted syllable is incorporated into a neighboring foot. Interestingly, an unpredicted pattern (**T2T2T3**)(**T2T3**) was attested across all age groups (under 10%), which we will return to discuss.

Our child and adult results allow us to reject **Binary dominance (H4)** since the larger domain pattern was the dominant pattern across groups, not the *L-to-R-Bin-Incorp* pattern. Also, children produced fewer larger domain patterns than adults in the five-digit sequences, and the use of this pattern increases with age (3-year-olds: 16.67%, 5-year-olds: 61.11%, and adults: 70%). In addition, at age 3, children's correct rates for the test items were between 20% - 30%, and at age 5, their correct rates were at about 60% - 70%. These results roughly translate to an increase in the correct rate by 40% in children's T3S application in flat structures in two years' time. These findings lend support to **Prosodic and T3S development (H5)**. Five-year-olds are still in the process of mastering T3S and still do not have adult-like performance, which leads us to the discussion of children's T3S errors.

The most common error type for 3-year-olds is over-application (e.g. ***T2T2**, ***T2T2T2**, ***T2T2T2T2T2**), indicating that although they have the knowledge of changing a T3 to a **T2** when followed by another T3, they have difficulty maintaining the underlying tone for the rightmost digit. The over-application can be over-generalization of the T3S rule, or it can be due to children's difficulty in production. After a few T2 it is easier to repeat one more T2. Five-year-olds had a relatively smaller proportion of such error type, but another common error type emerged—***T2T3T2** and ***T2T3T2T3T2**. This is a rather interesting finding that strongly indicates that they are using binary foot-building. These errors can be accounted for if the "incorporated" element was removed from "*L-to-R-Bin-Incorp*." Maintaining the same tone in a sequence (e.g. ***T2T2T2T2T2** or ***T3T3T3T3T3**) may be easier for 3-year-olds. Alternatively, but not as likely, three-year-olds might not have noticed the alternations between **T2** and T3 in binary feet (when a binary parsing strategy is used) that 5-year-olds have noticed. It should be emphasized that ***T2T3T2T3T2** is not the only possible error pattern that has alternations between **T2** and T3. ***T3T2T3T2T3** is also a possible error pattern that alternates between **T2** and T3. Nevertheless, it never surfaces. The absence of this pattern provides indirect evidence for left-to-right parsing.

Lastly, the pattern (**T2T2T3**)(**T2T3**) is not predicted by the T3S models [1-3, 9-10], and the pattern cannot be accounted for by either left-to-right or right-to-left parsing (predicting (**T2T3**)(**T2T2T3**) and (**T3T2T3**)(**T2T3**) respectively). It is not clear why adults also produced this pattern. A possible explanation is that both binary and ternary feet are available in flat structures. Namely, the five-syllable string is divided into a binary foot and a ternary foot, and both of which are available before the first syllable is produced. Either (**T2T3**)(**T2T2T3**) or (**T2T2T3**)(**T2T3**) surfaces, depending on whether a binary or ternary foot is selected first. A binary foot may not necessarily always precede other types of feet in digit parsing. Groupings of 2, 3, or 4 digits are common in phone numbers, social security numbers, and credit card numbers. The ternary foot in (**T2T2T3**)(**T2T3**) possibly occurs as a result of some digit-grouping strategy.

5. Conclusions

The predicted patterns for two- and three-digit sequences matched the patterns produced by children and adults. In a five-digit sequence, adults (but not children) produced the pattern (**T2T3**)(**T2T2T3**) predicted by *L-to-R-Bin-Incorp*. The larger domain pattern (**T2T2T2T2T3**) was the dominant pattern found in all age groups, and its use increases with age.

Both predicted parsing patterns are attested in the adult data. The unpredicted pattern (**T2T2T3**)(**T2T3**) may result from digit-grouping strategies. We suggest that foot-building in digits does not always follow the typical foot-building process. T3S in longer digit-sequences (e.g. 7, 9, 11 digits) can be further investigated to test the use of a combination of binary and ternary feet. Alternatively, repeated monosyllabic words (flat structures) can be tested in future studies to control for the factor of the possible digit-grouping in digit-sequences.

Regarding children's acquisition of T3S, 3- and 5-year-olds can apply T3S non-cyclically in digit sequences, although under- and over-applications are common error types. There is a clear developmental pattern in children's acquisition of T3S in flat structures as their parsing strategies develop with time. Five-year-olds' distinct error patterns, ***T2T3T2** and ***T2T3T2T3T2**, suggest that at age 5 (but not age 3), children are aware of the binary parsing strategy, although they have not fully acquired the incorporation of an unparsed syllable.

6. Acknowledgements

This research is funded by National Science Foundation under grant number BCS-0820614 (Doctoral Dissertation Research Improvement Grant). We also thank the participants of this study, and the reviewers for their invaluable comments.

7. References

- [1] M. Y. Chen, *Tone sandhi: Patterns across Chinese dialects*. Cambridge: Cambridge University Press, 2000.
- [2] C. Shih, "The prosodic domain of tone sandhi in Chinese," University of California San Diego PhD dissertation, La Jolla, CA, 1986.
- [3] C. Shih, "Mandarin third tone sandhi and prosodic structure," in *Studies in Chinese phonology*, N. Smith & J. Wang eds, Berlin: Mouton de Gruyter, 1997, pp. 81-124.
- [4] H.-H. Jeng, "The acquisition of Chinese phonology in relation to Jakobson's laws of irreversible solidarity," in *Proceedings of the 9th International Congress of Phonetic Sciences*, Copenhagen, Denmark, 1979, pp. 155-161.
- [5] H.-H. Jeng, "A developmentalist view of child phonology," *Studies in Language and Literature*, vol. 1, pp. 1-29, 1985.
- [6] C. Li and S. Thompson, "The acquisition of tone in Mandarin-speaking children," *Journal of Child Language*, vol. 4, pp. 185-199, 1977.
- [7] H. Zhu, *Phonological development in specific contexts: Studies of Chinese-speaking children*. Clevedon, UK: Multilingual Matters Ltd, 2002.
- [8] H. Zhu and B. Dodd, "The phonological acquisition of Putonghua (modern standard Chinese)," *Journal of Child Language*, vol. 27, pp. 3-42, 2000.
- [9] S. Duanmu, *The Phonology of Standard Chinese*, 1st/2nd ed. Oxford: Oxford University Press, 2000/2007.
- [10] S. Duanmu, "Left-headed feet and phrasal stress in Chinese," *Cahiers de linguistique - Asie orientale*, pp. 65-103, 2004.
- [11] Y.-H. Lin, *The sounds of Chinese*. Cambridge: Cambridge University Press, 2007.
- [12] Y.-c. Kuo, et al., "The phonetics and phonology of apparent cases of iterative tonal change in Standard Chinese," *Tones and tunes*, vol. 2, pp. 211-237, 2007.
- [13] D. McDaniel, et al., "Children's sentence planning: Syntactic correlates of fluency variations," *Journal of Child Language*, vol. 37, pp. 59-94, 2010.
- [14] S. Crain and R. Thornton, *Investigations in Universal Grammar: A guide to experiments on the acquisition of syntax and semantics*. Cambridge, MA: The MIT Press, 2000.
- [15] D. McDaniel, et al., *Methods for assessing children's syntax*. Cambridge, MA: The MIT Press, 1998.