INFANTS' RECOGNITION OF FUNCTION WORDS IN CONTINUOUS SPEECH

Rushen Shi

Département de psychologie, Université du Québec à Montréal
shi.rushen@uqam.ca

ABSTRACT

This study tests the hypothesis that function words are among the earliest word forms segmented by preverbal infants. In a visual fixation procedure, French-learning 8-month-old infants were familiarized to a function word, *mes* or *ta*. All infants were then tested with passages containing *mes* vs. *ta*. Looking times during the presentation of the two passage types were expected to differ if infants segmented the target functor. The results showed a significant interaction of passage type and sex. Although the direction of the looking preference is different for the two sexes, both groups showed a significant difference in listening times to the passage containing the target versus that containing the non-target. This suggests that both groups segmented the function words. The implications of functional elements for early lexical and syntactic acquisition are discussed. (Research funded by NSERC, SSHRC and CFI.)

Keywords: infant speech perception, function words, segmentation, lexical & syntactic learning

1. INTRODUCTION

Around one year of age infants begin to build a meaningful lexicon, and shortly afterwards they show evidence of comprehending some sentences e.g., [1]. This suggests that prior to these abilities infants have already succeeded in segmenting some word forms and larger linguistic units from continuous speech. In fact, given that parental speech contains mostly multi-word utterances, e.g., [2], segmenting speech into cohesive chunks is a prerequisite for lexical and syntactic acquisition.

Existing evidence indicates that infants begin to segment words and larger syntactic units from about 6 months of age. Infants use prosodic cues to segment clausal and phrasal units [3,4], and use several types of information to segment words, such as prosodic, phonotactic, statistical and coarticulation cues, e.g., [5-7]. Word segmentation studies have focused on content words, mostly nouns. Recent studies suggest that although function words emerge much later than content words in children's speech, they are among the earliest segmented and stored word forms in infants from 6 to 13 months of age [8-11]. Furthermore, function words enable infants to segment adjacent novel content words [12-14]. Function words also play a role in 11-month-olds' recognition of known words [15] and older infants' online language comprehension, e.g., [16,17].

The theoretical importance of functional morphemes has been proposed in language acquisition research [18-20]. According to this view, functional morphemes along with prosody can be used to bootstrap early lexical and syntactic acquisition, in contrast to the position that content words are primary in early language acquisition.

The role of functional morphemes for early lexical and syntactic learning hinges on infants' ability to segment functors from continuous speech. The present experiment tests if French-learning infants can segment function words from highly variable sentences with functors occurring in various utterance positions.

2. METHOD

Given that English-, French- and German-learning 8-month-olds can use function words to extract adjacent novel nouns [12-14], infants must already be able to segment and store function words at this age. Our experiment tests this hypothesis directly.

2.1. Participants

Eighteen Quebec-French-learning 8-month-olds completed this experiment. The data of four other infants were excluded due to fussiness (2), technical error (1), and parental interference (1).

2.2. Auditory stimuli

Two function words in French were chosen, *mes* (/me/, "my", plural, masculine and feminine) and *ta* (/ta/, "your", singular, feminine). Based on a spoken Quebec French database [21], the

frequency for *mes* is 1331, and for *ta* 613, much higher than the mean frequency of all words: 8.83.

In addition, two passages were created, one containing *mes* and the other *ta*. The passages were identical except for the target functors *mes* vs. *ta*. Each passage contained 6 sentences. Two of the sentences contained the target in sentence initial positions, in subject NPs. The other 4 sentences contained the target in sentence medial positions, with the functor heading different NPs of varying complexity. The sentences were each five to seven words long, containing seven to nine syllables.

A native Quebec-French speaker recorded the two passages and then the isolated function words. The stimuli were carefully chosen so that the target functors were comparable in overall prosodic variability, as were the two passages. Final stimuli contained one version for each passage and 18 citation tokens of each functor.

2.3. Design and procedure

The experiment was a direct segmentation task. Participants were randomly assigned to the *mes* familiarization condition or the *ta* familiarization condition, with sexes counter-balanced across infants. Each Familiarization trial presented the 18 tokens of a target functor, and the trial was repeated until the infant accumulated 30 sec of looking time. After familiarization, all infants were tested with two trial types: one with the *mes* passage, and the other with the *ta* passage. The *mes* passage type alternated with the *ta* passage type repeatedly during the Test phase, for a total of 5 trials for each type. The first Test trial was the *mes* passage or *ta* passage, counter-balanced across infants. The length for all trials was each 16.5 sec.

In a sound attenuated room, the infant sat on the parent's lap facing a central monitor. The parent heard masking music from headphones. The experimental software presented auditory stimuli and a black-and-white checkerboard simultaneously for each trial, and recorded all looks online. It also calculated all looks and moved the experiment from the Familiarization phase to the Test phase automatically. Each trial was initiated by the infant; that is, the researcher started a trial when the infant looked towards the monitor. Once started, the trial continued for the entire trial length. A red light flashed between trials to attract the infant's attention. The researcher, blind to the stimuli, observed the infant's eye responses through a closed-circuit TV and pressed a computer key whenever a look to the monitor occurred. A pre-trial presenting sinewave analogues of a non-word *neem* with the checkerboard served to acquaint the infant with the procedure. The sound was distinct from the experimental stimuli and would not interfere with the experiment.

All experimental sessions were offline coded for precision by another blind observer. The offline coded data were used for statistical analyses.

3. RESULTS

We calculated each infant's total looking time of all trials of each passage type during the Test phase, and then analyzed the overall response pattern across infants. We expected that if infants can segment functors from sentences, their looking time while listening to the target passage (i.e., containing the familiarized function word) should be different from their looking time while listening to the non-target passage (i.e., containing the non-familiarized function word).

The results showed that infants' looking times for the two passage types (target passage vs. non-target passage) were not different, $t(17)= -.406$; $p>.6$, 2-tailed. Mixed ANOVAs showed neither an interaction of Passage Type (within-subject factor) and Familiarization Target (between-subject factor: *mes* vs. *ta*), nor an interaction of Passage Type and Test Order (between-subject factor: target passage initial vs. non-target passage initial). No significant main effect was found. However, the results showed a highly significant interaction of Passage Type and Sex, $F(1,16)= 20.199$; $p< .0001$, despite the lack of a main effect of Passage Type.

Following this significant interaction, followup t-tests compared the looking times for the two passage types within each sex. For the boys, looking times were significantly different ($t(9)= -3.893$; $p=.004$, 2-tailed; target passage: Mean=53.67 sec, SE=5.17 sec; non-target passage: Mean=58.86 sec, SE=5.15 sec). The girls also exhibited a significant difference in looking times for the two passage types ($t(7)=2.609$; $p=.035$, 2-tailed; target passage: Mean=56.02 sec, SE=2.92 sec; non-target passage: Mean=51.04 sec, SE=4.2 sec). Overall, the boys preferred to listen to the non-target passage whereas the girls preferred the target passage. The two sexes showed opposite directions of looking preference. Thus, the earlier analyses with the two sexes combined masked the differential looking times for the two passage

types. The differential looking times for target vs. non-target passages suggest that both groups of infants segmented function words from sentences.

4. DISCUSSION

The present study shows that eight-month-old French-learning infants can segment function words from continuous speech. The sentences used in this experiment contained function words in medial positions in two thirds of the cases. Previous work showed that infants have more difficulty segmenting words from utterance-medial positions than utterance-edge positions [22], possibly due to coarticulation from the preceding and following speech contexts. Furthermore, function words are generally short and prosodically weak in comparison to content words [23]. It is therefore remarkable that the infants in our experiment segmented function words at the same age when content words are first segmented [5]. Our results are consistent with the existing findings that German and English preverbal infants perceive function words from sentences [10,11].

In our experiment boys and girls both showed a differential looking pattern during Test, although the directions of their preferences were not the same. The boys produced a novelty preference and the girls a familiarity preference. This appears intriguing. However, in existing word segmentation studies both familiarity preference and novelty preference are commonly observed, e.g., [5,6]. Artificial language tasks tend to yield novelty preferences whereas segmentation tasks using natural speech tend to produce familiarity preferences. The reasons for these differences are not well understood. In general, if an overall difference is observed for a group of infants, it is interpreted as evidence of word segmentation. In our study infants of the two sexes each showed a strong differential looking pattern, suggesting that they segmented the familiarized target functor from the Test sentences. The sex difference may be due to particular ways in which boys versus girls encode word forms under certain task conditions. In another study [8] in which we simplified the Test stimuli by using noun phrases instead of sentences, 8-month-olds segmented function words, and no sex difference was found. Thus, we are certain that preverbal infants can segment functors. The exact nature of different types of preferences and processing differences between boys and girls have received little attention in the field and need to be examined in future studies.

What are the mechanisms underlying infants' ability to segment function words? We suggest that the frequent occurrence of functional morphemes and their distinct prosodic properties are the likely language-universal factors for preverbal infants' recognition of functors. Firstly, across human languages, functional morphemes are extremely frequent relative to content words in adult-directed usage [21,24] and infant-directed speech, e.g., [23]. Infants are highly sensitive to statistics of the input speech, as shown in well-controlled artificial language tasks, e.g. [6]. It is thus plausible that the use of frequency cue is a determining factor underlying infants' early perception of functors. This was confirmed in our recent work on the role of frequency in English- and French-learning infants' perception of function words [12,13].

Secondly, functors are phonologically distinct from content words in infants' input, as shown in analyses of parental speech across typologically distinct languages [23]. Function words tend to be monosyllabic [23]. Monosyllabic words are more easily segmented by infants than bi- and multi-syllabic words, e.g., [25]. Functors often occur at major syntactic boundaries, preceded or followed by salient prosodic cues, making them potentially easier to segment. Function words are acoustically reduced across languages [23], so these words together with the adjacent, acoustically fuller content words form prosodic alternations. For example, NPs in English, French and many other languages typically begin with a determiner and are marked by a reduced-full prosodic alternation (e.g., *the dog; le chien*). The alternation is reversed in languages with the functor following the noun. Despite specific phonological differences between languages, such prosodic alternations are likely language general, thus may be salient to infants, especially when occurring at utterance edges. In sum, the frequency and prosodic characteristics may make functors accessible to infants at an early age. More work is needed to determine the relative importance of frequency and prosody in infants' recognition of functional morphemes.

The early recognition of functors may directly impact infants' lexical and syntactic acquisition [18-20]. Unlike frameworks that place importance on content words for initial language acquisition, function words may be essential for bootstrapping early language learning tasks. For instance, infants

may use functors to segment novel words [12-14], assign words to appropriate grammatical classes [26,27], parse larger syntactic constituents, distinguish and labeling different syntactic constituents, etc.. Segmenting and storing the forms of functional morphemes are necessary for their subsequent roles in language acquisition. Our study provides evidence that this process indeed emerges at the preverbal stage, before the onset of vocabulary and syntactic learning.

5. REFERENCES

[1] Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J. 1994. Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59* (Serial No. 242, No. 5).

[2] Van de Weijer, J. 1998. *Language input for word discovery*. Ph.D. dissertation, University of Nijmegen.

[3] Nazzi, T., Kemler Nelson, D., Jusczyk, P., & Jusczyk, A.M. 2000. Six-month-olds' detection of clauses in continuous speech: Effects of prosodic well-formedness. *Infancy*, 1, 123-147.

[4] Soderstrom, M., Seidl, A., Kemler Nelson, D., Jusczyk, P. 2003. The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *J. Mem..Lang., 49*, 249-267.

[5] Jusczyk, P., Houston, D., Newsome, M. 1999. The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39*, 159-207.

[6] Saffran, J., Aslin, R., Newport, E. 1996. Statistical learning by 8-month-old children. *Science, 274*, 1926-28.

[7] Curtin, S., Mintz, T. Byrd, D. 2001. Coarticulatory cues enhance infants' recognition of syllable sequences in speech. *Proc. 25th BUCLD*, Boston.

[8] Shi, R. Marquis, A., Gauthier, B. 2006. "Segmentation and representation of function words in preverbal French-learning infants." *Proc. 30th BUCLD*, Boston..

[9] Shi, R., Werker, J., Cutler, A. 2006. Recognition and representation of function words in English-learning infants. *Infancy, 10*, 187-198.

[10] Höhle, B., Weissenborn, J. 2003. German learning infants' ability to detect unstressed closed class elements in continuous speech. *Dev. Sci., 6*, 122-127.

[11] Shady, M. 1996. *Infants' sensitivity to function morphemes*. Doctoral dissertation, SUNY Buffalo.

[12] Shi, R., Cutler, A., Werker, J., Cruickshank, M. 2006. Frequency and form as determinants of functor sensitivity in English-acquiring infants. *J. Acoust. Soc. Am. 119*, EL61-EL66.

[13] Shi, R., Lepage, M. In press. The effect of functional morphemes on word segmentation in preverbal infants. *Dev. Sci.*

[14] Höhle, B., Weissenborn, J. 2000. The origins of syntactic knowledge: Recognition of determiners in one year old German children. *Proc. 24th BUCLD*, Boston.

[15] Hallé, P., Durand, C., de Boysson-Bardies, B. Submitted. Do 11-month-old French infants process articles?

[16] Johnson, E. 2005. Grammatical Gender and Early Word Recognition in Dutch. *Proc. 29th BUCLD*, Boston.

[17] Gerken, L.A., Landau, B., Remez, R. 1990. Function morphemes in young children's speech perception and production. *Dev. Psych., 27*, 204-216.

[18] Christophe, A., Guasti, M. T., Nespor, M., Dupoux, E., van Ooyen, B. 1997. Reflections on prosodic bootstrapping: its role for lexical and syntactic acquisition. *Lang. & Cog. Processes*, 12, 585-612.

[19] Morgan, J., Shi, R., Allopenna, P. 1996. Perceptual bases of rudimentary grammatical categories. In J. Morgan & K. Demuth (Eds.), *Signal to syntax* (pp. 263–283). Hillsdale, NJ: Lawrence Erlbaum Associates.

[20] Shi, R. 2005. Early syntactic categories in infants. In Cohen, Henri & Claire Lefebvre, (eds.), *Handbook of categorisation in cognitive science*. Elsevier.

[21] Beauchemin, N., Martel P., Théoret, M. 1992. *Dictionnaire de fréquence des mots du francais parlé au Québec : Fréquence, dispersion, usage, écart réduit.* New York: Peter Lang Publishing.

[22] Seidl, A., Johnson, E. 2006. Infant word segmentation revisited: edge alignment facilitates target extraction. *Dev. Sci., 9*, 565-573.

[23] Shi, R., Morgan, J. L., Allopenna, P. 1998. Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *J. Child Lang., 25*, 169-201.

[24] Kucera, H., Francis, W. N. 1967. *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

[25] Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., Alcantara, C. 2006. Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *J. Mem.Lang., 54*, 283-299.

[26] Höhle, B. Weissenborn, J., Kiefer, D., Schulz, A. Schmitz M. 2004. Functional elements in infants' speech processing: The role of determiners in the syntactic categorization of lexical elements. *Infancy* 5(3), 341-353.

[27] Mintz, T. 2006. Finding the verbs: distributional cues to categories available to young learners. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action Meets Word: How Children Learn Verbs*, p. 31-63. NY: Oxford Uni. Press.

6. APPENDICES

A: Average acoustic values (& SD) of vowels of function words across tokens during Familiarization and Test

	Duration (ms)	Mean F0 (Hz)	Mean Amp (db)
mes (18 tokens; citation; Fam)	270.0 (36.96)	243.94 (66.68)	65.5 (2.68)
ta (18 tokens; citation; Fam)	256.17 (41.35)	237.78 (71.47)	67.22 (2.21)
mes (6 tokens; in sentences; Test)	108.83 (22.11)	184.67 (4.93)	67.17 (3.19)
ta (6 tokens; in sentences; Test)	113 (23.71)	183.33 (11.2)	66 (1.67)

B: Passages during the Test phase

La chance épargne *ta* ville miniature. Je déchire toujours *ta* jupe. *Ta* boulangère brûle le pain. Un virus plonge souvent dans *ta* joue. *Ta* fillette renverse le verre d'eau. Des roches percent *ta* chaussette de soie.

La chance épargne *mes* villes miniatures. Je déchire toujours *mes* jupes. *Mes* boulangères brûlent le pain. Un virus plonge souvent dans *mes* joues. *Mes* fillettes renversent le verre d'eau. Des roches percent *mes* chaussettes de soie.