Segmentation and Representation of Function Words in Preverbal French-Learning Infants

Rushen Shi, Alexandra Marquis, and Bruno Gauthier
Université du Québec à Montréal

1. Introduction

Functional elements are crucial for the structure of the grammar. Both innatist and inductionist frameworks of language acquisition pay particular attention to the development of these elements. Classic work showed that function words and functional morphemes start to emerge in children’s speech much later than content words; children’s early production is typically telegraphic (e.g., Brown, 1973). Recent research using more sensitive tasks, however, has revealed a different picture: children begin to analyze functional elements before they could reliably produce them. Gerken and colleagues (Gerken, Landau, & Remez, 1990; Gerken & McIntosh, 1993) demonstrated that two-year-old children, who generally omit functional items in their speech, were nonetheless sensitive to functional morphemes in sentence imitation and picture-naming tasks; their performance was enhanced by the presence of appropriate functors. These results are consistent with the argument that function words may play an important role in lexical and syntactic acquisition (e.g., Christophe, Guasti, Nespor, Dupoux, & van Ooyen, 1997; Morgan, 1986).

More recent studies measuring the time-course of online language comprehension showed that English-learning infants at 18 months of age process determiners in the interpretation of the adjacent nouns (Zangl & Fernald, 2005), and Dutch-learning 2-year-olds encode gender specific information in determiners associated with the adjacent nouns (Johnson, 2004). Moreover, early in the second year of life infants are sensitive to co-dependencies of non-adjacent functional elements in sentences (e.g., Santelmann & Jusczyk, 1998) and derive syntactic categories (e.g., nouns, verbs) based on the distribution of co-occurring functional elements (e.g., Mintz, in press; Höhle, Weissenborn, Kiefer, Schulz, & Schmitz, 2004). These findings suggest that infants must have already successfully analyzed the spoken forms of functional elements and have stored them in memory before these elements exert effects on the interpretation of adjacent content words and on the processing of the sentences.

The possibility that functional elements are processed early in infancy, before the emergence of the first words, is a reasonable one. Functional items are small in the total number of types, but each occurs extremely frequently. Content words, in contrast, are large in the number of types, but each in general occurs far less frequently. This frequency distinction appears in infant-directed
speech across typologically different languages (Shi, Morgan, & Alloppenna, 1998), as well as in adult spoken language (Cutler & Carter, 1987). This asymmetry of frequencies has been shown to be critical for deriving a grammar in adult experiments (Valian & Coulson, 1988). Infant speech perception research has shown that during the second half of the first year infants can use frequency/distributional cues to learn about the phonetic inventory and phonotactic patterns of their native language (Jusczyk, Charles-Luce, & Luce, 1994; Maye, Werker, & Gerken, 2002), and to segment word forms in tasks using artificial languages (Saffran, Aslin, & Newport 1996). It is thus plausible that infants, who hear functional elements so frequently in various contexts, may extract and store these words before they show first evidence of language comprehension and production.

Studies with preverbal infants indeed have shown evidence of early processing of functional elements. English-learning infants from 10.5 to 13 months of age detect real functors with varying frequencies from nonsense functors (Shady, 1996; Shafer, Shucard, Shucard, & Gerken, 1998), even when the nonsense functors were only minimally modified from real functors (Shi, Werker, & Cutler, in press). At eight months English infants can recognize high-frequency functors (Shi, Cutler, Werker, & Cruickshank, submitted).

In the present study we ask when French-learning preverbal infants begin to segment functors from continuous speech and whether frequency is a determining factor. French functors such as determiners occur consistently in their syntactic position within noun phrases. Thus they may be segmented very early in infancy. Moreover, functors in French are less reduced than in English, and may be more perceptually salient to younger infants. Likewise, functors in German are also prosodically stronger than those in English. Indeed, they are segmented as early as seven months of age (Höhle & Weissenborn, 2003).

The second goal of this study was to examine the nature of phonetic encoding of function words in French-learning infants. Precise representations of functional elements are crucial for the acquisition of syntactic structures, as will be discussed further in General Discussion. To date, research on functor encoding has only focused on English-learning infants, who showed an initial underspecification at eight months of age and detailed representation by 11 months of age (Shi, et al., in press; Shi, et al., submitted). The lack of specification may be due to the especially reduced forms of functors in English (Cutler, 1993). The question is whether functors in a syllable-timing language such as French also exhibit the same developmental pattern.

2. Experiment 1

In a recent study we showed that French-learning eight-month-old infants can segment function words (Shi & Gauthier, 2005). Infants were familiarized with a target determiner and tested on phrases containing that target versus phrases containing a different determiner. Infants looked significantly longer while listening to the phrases containing the non-familiarized determiner than
those containing the target. This novelty effect is consistent with the existing view that when a representation is first constructed, infants tend to show a familiarity preference, but as the representation becomes more robust, a novelty preference is then observed (e.g., Roder, Bushnell, & Sasseville, 2000). In the present experiment we predict that infants may begin to segment function words at an age younger than eight months, by producing a familiarity preference. Previous work showed that infants from 4.5 months begin to segment very familiar words, such as their own names, mommy, and daddy (e.g., Mandel, Jusczyk, & Pisoni, 1995), which are likely frequent words that infants hear. Moreover, French determiners are highly consistent in occurrence within a noun phrase. French has both definite and indefinite articles, which must occur in the same syntactic position (e.g., *la table* “the table”, *des tables* “tables”). Whereas common nouns in isolation can occur in English (e.g., dogs), such expressions are ungrammatical in French. The regular distribution and the more syllabic nature of functors in French (compared to English) may be highly salient to young infants.

2.1.1. Participants and Stimuli

Sixteen six-month-old infants completed this experiment. Infants were monolingual Quebec French learners. Stimuli were the same as those in Shi and Gauthier (2005). They were the definite article *la* (/la/, feminine, singular), and the indefinite article *des* (/de/, masculine and feminine, plural), each occurring in different noun phrases, i.e., *la sangle, la blouse, la guilde, la preuve, des sangles, des blouses, des guildes, des preuves*. The frequencies of the four nouns are rare in adult speech in French (Beauchemin, Martel, & Théoret, 1992). Infants were unlikely to have had previous exposure to them, so that any effect shown in the experiment could be interpreted as the effect of the functors. Two of the nouns included onset consonant clusters, and two included coda consonant clusters. Note that the *s* in *des* and at the end of all nouns is not pronounced in French. The syllabic structure of the nouns was more complex and thus distinct from the typical forms of functors. A native speaker of Quebec French, a mother of a preverbal infant, recorded the noun phrases and then the determiners in isolation in an infant-directed style. She was not informed of the goal of the study. The final stimuli set consists of multiple tokens of each noun phrase, as well as multiple tokens of *la* and *des* in isolation.

Since wave analogues of a non-word *neem* were used as pre- and post-test stimuli (Vouloumanos, Kiehl, Werker, & Liddle, 2001). The pre-test served to introduce the infant to the task. The infant was expected to recover in looking time during the post-test, allowing us to determine if infants were on task.

2.1.2. Design

As in Shi and Gauthier (2005), we used a two-phase (Familiarization-Test) task introduced by Jusczyk and Aslin (1995) for testing infant speech
segmentation. Infants were assigned randomly to the *la* condition or *des* condition. They were first familiarized to multiple tokens of the target functor (*la* or *des*) produced in isolation. Each trial was 16.5 seconds long. After 30 seconds of looking time was accumulated, the Familiarization Phase ended, followed immediately by the Test Phase. Each infant heard two types of Test trials presented in alternation, one with the noun phrases containing *la* (e.g., *la sangle*), another with the noun phrases containing *des* (e.g., *des sangles*). The noun phrases within each trial occurred randomly. There were a total of 10 Test trials, five of each type. The trial length was 18.5 seconds long. The first Test trial was either that of *la*-phrases or *des*-phrases, counterbalanced across infants. Inter-phrase duration was 600 ms, with minor adjustments to maintain the same trial length, despite small durational variations for different phrases.

### 2.1.3. Procedure

The infant was seated on the parent’s lap facing a TV monitor and a loudspeaker in a sound-attenuated chamber. The parent wore headphones to hear masking music. Using the Habit software (Cohen, Atkinson, & Chaput, 2000), we measured infants’ visual responses to the central visual display of a black-and-white checkerboard that was simultaneously presented with the auditory stimuli from the central speaker. The researcher, blind to the target word and the order of presentation, observed the infant’s eye movements from an adjacent room through a closed-circuit TV monitor, and pressed a computer key as soon as a look to the center was delivered by the infant. Each trial was initiated upon the infant’s fixation to the monitor, and then stayed on until the end of the trial. The software recorded all looks during a trial and calculated online the total looking time for each trial. All test sessions were videotaped, so that offline frame-by-frame coding could be performed to verify the online coding.

### 2.2. Results

For each infant, mean total looking times across the five Test trials containing the familiarized target functor and that across the five Test trials containing the non-familiarized functor were calculated. They were then compared in a 2x2 mixed ANOVA, with Functor Type (familiarized vs. non-familiarized) as the within-subject factor, and Target (*la* vs. *des*) as the between-subject factor. The results reveal a significant main effect of Functor Type, $F(1, 14) = 6.824, p = .02$. Unlike the eight-month-olds in our previous study, who showed a novelty preference, the six-month-old infants in this experiment produced a familiarity preference. They looked longer while listening to phrases containing the familiarized target functor than while listening to those containing the non-familiarized functor. There was no main effect of Target, nor Target x Functor Type interaction. These results, combined with those of Shi and Gauthier (2005), suggest that French-learning infants at six months of age begin to segment *des* and *la*, and this ability is more robust by eight months.
Fig. 1. Six-month-old infants’ segmentation of des and la.

3. Experiment 2

Experiment 1 established that six-month-old French-learning infants begin to segment function words from connected speech. The functors la (/la/) and des (/de/) were phonetically distinct from each other, differing in both the syllable onset consonants and the vowels. In Experiment 2 we examined further how well six-month-old infants encode the phonetic forms of the segmented functors. To do so, we used target functors that are more phonetically similar.

3.1.1. Participants, Stimuli, Design and Procedure

Sixteen six-month-old Quebec-French-learning infants completed the task. The functor la and the noun phrases containing la created for Experiment 1 were used. In addition, a phonetically similar functor ta (i.e., /ta/, “your”, singular, feminine) was included. Note that the initial /t/ in French is not aspirated, thus more similar to /l/ than /t/-/l/ distinction in English. The ta tokens and the noun phrases containing ta were recorded by the same speaker, at the same time when the stimuli for Experiment 1 were recorded. Therefore, the stimuli set for Experiment 2 included isolated tokens of la and ta, and phrases containing them (i.e., la sangle, la blouse, la guilde, la preuve, ta sangle, ta blouse, ta guilde, ta preuve). The sine wave analogues of speech for pre- and post-tests were the same as in Experiment 1. The design of Experiment 2 was identical to that of Experiment 1, except that the isolated ta tokens and the noun phrases containing ta were used instead of the isolated des tokens and des phrases. The procedure was as in Experiment 1.

3.2. Results

As in Experiment 1, infants’ total looking times during the Test trials were analyzed in a 2x2 mixed ANOVA, with Functor Type (familiarized vs. non-
familiarized) as the within-subject factor, and Target (la vs. ta) as the between-subject factor. There was no main effect of Functor Type, nor Target, nor Functor Type x Target interaction. Recall that Experiment 1 showed successful segmentation of la when it was tested against des. The combined results of Experiment 1 and 2 suggest that infants did not distinguish between la and ta in a segmentation task. Although they can extract the word forms of functors at six months of age, the functors are not encoded with well specified phonetic representations.

Fig. 2. Six-month-old infants’ segmentation of la and ta.

However, could the null results of Experiment 2 be simply due to a lack of the ability to distinguish the consonants in la and ta in general, rather than revealing anything about word forms? This is unlikely, given the existing evidence of infants’ remarkable ability to distinguish phonetic contrasts very early in infancy. Experiment 3 is designed to verify this possibility.

4. Experiment 3

In this experiment we examined if infants can discriminate la from ta in a habituation-dishabituation task. If the null effect for the target functors in Experiment 2 was indeed due to the failure to encode the phonetic details of functor forms during word segmentation rather than due to an inability to distinguish /l/ and /t/, then infants should succeed in distinguishing the two words in a pure discrimination task.

4.1.1. Participants and Stimuli

Twenty-three six-month-old infants completed this experiment. Infants were monolingual Quebec French learners. The isolated tokens of la and ta used for Experiment 2 were used for Experiment 3.
4.1.2. Design and Procedure

In this task, infants were first habituated to trials of one of the function words. For example, one group of infants heard la trials, i.e., variable tokens of la, until they reached the pre-established habituation criterion, 65% decline in comparison to the first window of trials. The calculation was done online by the Habit software using three-trial sliding windows. Once the habituation was attained, the Test Phase started immediately, presenting a Same Trial (i.e., tokens of the same functor as in the Habituation Phase), and a Change Trial (i.e., tokens of a different functor). The Same Trial, however, was a different set of tokens of the same habituated functor. This manipulation ensures that the infants were performing a categorical discrimination task, rather than purely discriminating on the basis of memorizing the exact tokens of the Habituation set. Infants were assigned randomly to the la or ta Habituation condition, and to the Same first or Change first order. Thus, all conditions and orders were counterbalanced across infants. The procedure was as in Experiment 1.

4.2. Results

Infants’ looking times during the last Habituation trial and each of the two Test trials were analyzed in a 2 x3 mixed ANOVA, with Habituation Functor (la vs. ta) as the between-subject factor, and Trial Type (Last vs. Same vs. Change) as the within-subject factor. The results indicated a significant main effect of Trial Type, $F(2, 42)= 6.732, p= .003$, and no other main effect nor interaction was found. Planned comparisons showed a significant difference between the

![Fig. 3. Six-month-old infants’ discrimination of la vs. ta.](image)

Last and the Change trials, $F (1, 21)= 10.945, p= .003$, and between the Same and Change trials, $F (1, 21)= 7.222; p= .014$, with a longer looking time for the Change trial in both comparisons; in contrast, the Last and Same trials were not different ($F < 1$), even though the Same trial contained novel tokens of the
Habituation functor. These results suggest that infants discriminated *la* and *ta*, thus the null results in Experiment 2 indeed reflect six-month-old infants’ failure to encode phonetic details of the functors during word segmentation, and not an inability to distinguish the initial consonants in *la* and *ta*. In Experiment 4 we asked if by eight months of age infants are better able to encode the phonetic details of functors.

5. Experiment 4
5.1.1. Participants, Stimuli, Design and Procedure

Sixteen eight-month-old infants completed this experiment. Infants were monolingual Quebec French listeners. The stimuli and design were identical to those of Experiment 2. The procedure was as in Experiment 1.

5.2. Results

As in Experiment 1 and 2, infants’ mean total looking times were analyzed in the same 2 (Functor Type: familiarized vs. non-familiarized) x 2 (Target: *la* vs. *ta*) mixed ANOVA. There was no significant main effect of Target. The main effect of Functor Type was marginally significant, *F*(1, 14) = 3.866, *p* = .069. However, there was also a marginally significant Functor Type x Target interaction, *F*(1, 14) = 3.644, *p* = .077. Infants familiarized with *la* yielded significantly longer looking time while listening to Test trials containing *la* than those containing *ta* (paired *t*(7) = 2.915, *p* = .022, 2-tailed). In contrast, infants familiarized with *ta* did not show any difference while listening to the Test trials containing the target *ta* versus those containing the non-target *la*. (See Fig. 4.)

The results suggest that infants can encode the phonetic details of *la* during word segmentation by eight months of age, a progress from six months of age. However, *ta* showed no progress in eight-month-old infants during segmentation. This result is interesting. A plausible interpretation could be considered in terms of frequency. The functors *la* and *des* occur much more frequently than *ta*, although all of these are generally more frequent than content words. The frequency of *la* is 13312, but that of *ta* is only 613 according to the Beauchemin spoken Quebec French corpus (Beauchemin, et al., 1992). The same pattern holds in child-directed speech in French (e.g., Champaud corpus in CHILDES, MacWhinney, 2000). Thus, infants at eight months of age may not have heard enough instances of *ta* in their listening experience such that the word may not have yet been stored with phonetic details. In contrast, infants have had sufficient exposure to the highly frequent functor *la* by this age and have no difficulty encoding its details during word segmentation. Note that infants in the *la* and *ta* conditions received the same amount of Familiarization exposure. Thus, the differential result patterns for the Test trials across the two Target conditions must be related to infants’ prior experience with the two functors. This frequency effect is particularly striking given that /l/ is generally a more unmarked consonant in comparison to /t/.
8-Month-Old Infants’ Segmentation of la and ta

Fig. 4. Eight-month-old infants’ segmentation of la and ta.

It is interesting that whereas six-month-olds showed no differential looking time when tested in the la-ta segmentation task (Experiment 2), eight-month-olds produced a familiarity preference for the segmentation of la in this la-ta task. Note also that infants at eight months showed a novelty preference in a la-des segmentation task (Shi & Gauthier, 2005), an apparently easier task due to the larger phonetic differences between the functors la-des. The same la-des task yielded a familiarity preference in six-month-olds (Experiment 1), whose segmentation ability was just emerging. Together, these results reveal that infants have made progress in encoding functors from six to eight months.

6. General Discussion

The present study showed a clear developmental progression in French-learning infants’ recognition of functors from six to eight months of age. The segmentation of highly frequent function words in Quebec French emerges at six months of age, and becomes more robust by eight months. Segmented function words, however, are phonetically underspecified in younger infants. By eight months of age high-frequency function words are represented with full specification, whereas lower-frequency functors are still not encoded in details. Furthermore, we showed that six-month-old infants have no difficulty distinguishing phonetically similar function words in a pure discrimination task. Thus their performance in the word segmentation task reflects a failure in encoding the words while extracting them from connected speech.

A number of observations can be made on our results in relation to the existing findings of function word segmentation and representation. First, function words in Quebec French are shown here to be segmented at an earlier age than those in English (Shady, 1996; Shafer, et al., 1998; Shi, et al., in press), potentially due to the differences in rhythmic and distributional properties between the two languages. German function words, which are less reduced than English functors, are segmented by seven- to nine-month-olds, but not six-
month-olds (Höhle & Weissenborn, 2003). Our study reveals that this ability is present in infants learning Quebec French at six months.

Second, frequency is a determining factor in French-learning infants’ segmentation and phonetic encoding of function words, shown in this study and in previous studies with English-learning infants (Shi, et al., in press; Shi, et al., submitted). In both languages the most frequent function words are the first to be segmented and well specified in form.

Finally, the phonetic encoding of function words in Quebec French evolves from being underspecified to being fully specified, as in the case of English infants. Underspecification was observed for English functors in eight-month-old infants, and detailed encoding in 11-month-old infants (Shi, et al., in press; Shi, et al., submitted). In French-learning infants the progression follows the same pattern, although the time course is shifted earlier than English infants.

A great deal of attention has been given to the nature of phonetic encoding for content words (especially nouns) in infants. Detailed specifications of nouns have been shown in word segmentation experiments in infants as young as six and eight months of age (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Jusczyk & Aslin, 1995), and in word comprehension and word learning tasks in infants from one to two years of age (e.g., Swingley & Aslin, 2000; Swingley & Aslin, 2002; Werker, Fennell, Corcoran, & Stager 2002; but see Hallé & de Boysson-Bardies, 1996 and Vihman, Nakai, DePaolis, & Hallé, 2004 for evidence of underspecification). Encoding content words with phonetic details allows for contrasting many words, a requirement for building a large lexicon.

In contrast, detailed representations of functional elements are essential for the development of the grammar. Since the syllabic structure of functors is simpler than that of content words, the forms of many functors are similar, including a significant proportion of minimal pairs. Infants must learn to distinguish many phonetically similar forms of functors so as to acquire the syntactic relationships defined by various functors. Indeed, infants are performing such syntactic analysis early in the second year of life (e.g., Santelmann & Jusczyk, 1998). Fully specified functors are also important in their facilitative roles for the learning and processing of content words. For instance, syntactic category differences between the noun and verb in the dog and to eat may be derived by their co-occurrence patterns with the and to, which are typically produced with only a difference in the consonants in spontaneous speech. This distinction must be encoded if infants use them to learn the category membership or the word meanings of the adjacent content words.

Acquiring a grammar is a complex task for children. Functional elements may play a crucial role in this process. Extracting functors from continuous speech and encoding them in the memory with detailed representations are prerequisites for their role in language acquisition. Our experiments demonstrate that functional elements begin to be segmented by six months of age, and the phonetic representations of functors develop from being underspecified to being well specified within the first year of life, just in time to potentially assist infants in various language learning tasks.
Author notes

This research was funded by grants from NSERC, FQRSC, and SSHRC to the first author. We thank Mélanie Lepage and David-Étienne Bouchard for research assistance. Correspondence regarding this article should be addressed to Rushen Shi, Département de psychologie, Université du Québec à Montréal, Montréal, QC, H3C 3P8, Canada. (shi.rushen@uqam.ca)

References


Werker, J., Fennell, C., Corcoran, K., & Stager, C. (2002). Infants' ability to learn phonetically similar words: effects of age and vocabulary size. Infancy, 3(1), 1-30.