Syntactic Categorization in French-Learning Infants

Rushen Shi and Andréane Melançon
Département de Psychologie
Université du Québec à Montréal

Recent work showed that infants recognize and store function words starting from the age of 6–8 months. Using a visual fixation procedure, the present study tested whether French-learning 14-month-olds have the knowledge of syntactic categories of determiners and pronouns, respectively, and whether they can use these function words for categorizing novel words to nouns and verbs. The prosodic characteristics of novel words stimuli for noun versus verb uses were balanced. The only distinguishing cue was the preceding determiners versus subject pronouns, the former being the most common for nouns and the latter the most common for verbs, i.e., Det + Noun, Pron + Verb. We expected that noun categorization may be easier than verb categorization because the co-occurrence of determiners with nouns is more consistent than that of subject pronouns with verbs in French. The results showed that infants grouped the individual determiners as one common class, and that they appeared to use the determiners to categorize novel words into nouns. However, we found no evidence of verb categorization. Unlike determiners, pronouns were not perceived as a common syntactic class.

INTRODUCTION

An important task during language acquisition is the assignment of words to grammatical categories (e.g., nouns, verbs, determiners, and pronouns).

Correspondence should be sent to Rushen Shi, Département de Psychologie, Université du Québec à Montréal, C.P. 8888, Succursale Centre-Ville, Montréal, QC, Canada H3C 3P8. E-mail: shi.rushen@uqam.ca

This is because such categories are the basic building blocks of syntax. Syntactic structures underlying possible utterances in any language are defined in terms of grammatical categories. This characteristic reflects the productive nature of human languages which all learners acquire—a finite set of structures and principles expressed in syntactic categories can represent an infinite number of utterances. The grammatical category information, however, must be learned individually for each word. While adults learning a second language are often overtly given information about parts of speech for new vocabulary items, infants do not receive any explicit instruction of this type. In this study, we inquire how infants learn to assign words to their appropriate grammatical categories.

Different theoretical models have been proposed to understand the mechanisms underlying initial grammatical categorization. One type of theories focuses on the use of semantic cues for assigning words to initial grammatical categories (e.g., Braine, 1976; Pinker, 1984). Infants may construct initial grammatical categories by mapping words to conceptual categories, for example, by noticing that words denoting objects, words denoting actions, and words denoting properties belong to separate classes (e.g., Gentner, 1982). According to nativist views, infants come to the task with innate syntactic categories and innate linking rules that allow them to associate content words to appropriate grammatical categories through meaning (Pinker, 1984). This semantic-based classification can later be combined with distributional based analyses, and needs to be transformed into more abstract grammatical categories.

Another type of models stresses the direct analysis of the input without the mediation of semantic/conceptual categories. Infants may potentially break into the syntactic system by classifying words according to their differential phonological, acoustic, and distributional properties. Indeed, such cues supporting basic grammatical distinctions have been shown to exist in parental speech (e.g., Conwell & Morgan, 2009; Mintz, 2003; Monaghan, Chater, & Christiansen, 2005; Monaghan, Christiansen, & Chater, 2007; Shi & Moisan, 2008; Shi, Morgan, & Allopenna, 1998). It has also been proposed that learners can achieve initial grammatical categorization solely on distributional grounds (e.g., Maratsos & Chalkley, 1980), by formal analyses of the distributional patterns of the preceding and following contextual words (e.g., Mintz, 2003; Redington, Chater, & Finch, 1998). As grammatical categories are by definition abstract syntactic constructs in terms of their relations with one another in the phrase structure, this approach is more direct in addressing the learning problem, avoiding the need for the transition from semantic categories to syntactic categories.

There is evidence that infants shortly after 1 year of age can perform basic grammatical categorization in the absence of semantic cues. Infants

learning English have been shown to categorize verbs (Mintz, 2006), and German-learning infants showed the ability to categorize nouns (Höhle, Weissenborn, Kiefer, Schulz, & Schmitz, 2004). Infants in these studies demonstrated rudimentary understanding of the grammatical categories of novel words by noticing their context, for example, perceiving that nouns in German can co-occur with different determiners. In experiments using artificial languages (Gomez & Lakusta, 2004), infants aged 12 months categorized novel words based on a phonological feature shared among members of the same class and its association with specific adjacent items that resembled closed-class items. American infants could categorize Russian noun gender classes after being trained with exemplars exhibiting regular distributional patterns between the noun stems and gender morphemes, and the integration of double markings (both the gender ending following the noun stem and an additional gender-marking element at the end of the stem) was necessary (Gerken, Wilson, & Lewis, 2005). A recent study on word learning showed that 23-month-old French toddlers attend to function word cues to nouns and verbs, and can learn the meanings of novel verbs using the category information signaled by function words (Bernal, Lidz, Millotte, & Christophe, 2007). A common characteristic that these studies share is the occurrence of closed-class items as the contextual support for categorization, although other cues (e.g., prosodic) may also have existed in the novel word stimuli and affected infants' responses.

The idea that function words/morphemes may assist grammatical categorization of lexical words is reasonable in that these items are a small set that occur extremely frequently across human languages. It is known from studies using artificial languages that infants are capable of tracking frequency and other distributional information (e.g., transitional probabilities between elements, as shown in Saffran, Aslin, & Newport, 1996). In adult artificial language learning experiments, more frequent occurrence of a small number of closed-class items was important for assisting grammatical categorization of adjacent words (Valian & Coulson, 1988). Closed-class items also exhibit distinct phonological and acoustic characteristics compared with open-class words (e.g., Monaghan et al., 2005; Shi et al., 1998). The categorization of lexical versus function words based on their distinctive acoustic and phonological properties has been shown in newborn infants (Shi, Werker, & Morgan, 1999). After 6 months of age infants begin to store individual function words in the memory and use them to segment adjacent words (e.g., Hallé, Durand, & de Boysson-Bardies, 2008; Höhle & Weissenborn, 2000, 2003; Shady, 1996; Shafer, Shucard, Shucard, & Gerken, 1998; Shi, Cutler, Werker, & Cruickshank, 2006; Shi & Lepage, 2008; Shi, Marquis, & Gauthier, 2006; Shi, Werker, & Cutler, 2006). It is plausible that infants may subsequently divide

function words into smaller categories (e.g., determiners and pronouns) and use them to classify the adjacent words.

In the previous categorization study with German-learning infants (Höhle et al., 2004), nonwords were paired with a determiner or a pronoun during the familiarization phase. During the test phase, infants heard sentences containing the nonwords used as nouns in novel contexts versus sentences containing the nonwords used as verbs in novel contexts. Infants aged 14–16 months who were familiarized with the nonwords in the determiner context differentiated the two types of test sentences, whereas those who were familiarized with the pronoun context did not. These results indicated the ability in categorizing nouns but not verbs in German infants. The study of Mintz (2006) tested 12-month-old English-learning infants' categorization of nouns versus verbs. Mintz used sentences in both familiarization and test phases. Nonwords had both preceding and following contexts during familiarization. They were in different sentential contexts during the test phase. The infants in Mintz categorized verbs but not nouns. Höhle et al. (2004) and Mintz attributed their findings to the role of contextual function words, assuming that infants know some individual function words. What remains unknown is whether infants have the knowledge about function word subclasses, e.g., the determiner category and the pronoun category.

In this study, we seek to better understand infants' early grammatical categorization. First, we aimed at testing whether infants shortly after 1 year of age have the basic knowledge of specific functional categories, in particular, whether they know that individual determiners belong to one class and that individual pronouns belong to another class. These functional categories represent a crucial part of the human syntactic knowledge. Any empirical evidence indicating that infants have these categories at the earliest stage of syntactic acquisition would be welcomed by theories claiming that syntactic categories are innate. The previous infant categorization studies by Höhle et al. (2004) and Mintz (2006) used complex, varying sentential contexts during their test phase. They were not designed to test function word categorization. In the present study, we designed our stimuli differently, in that both familiarization and test phases included only the immediately preceding function words (determiners versus pronouns). The nonwords were kept constant throughout the experiment. As the function words during the test phase were new, with one conforming to and another not belonging to the category of the familiarization function words, we were able to directly assess whether infants formed basic groupings with the function words in the stimuli set, i.e., treated the individual determiners as one class and pronouns as a distinct class.

Second, we extended the previous work to French-learning infants, to determine whether noun and verb categorization can be driven by adjacent frequent function words in French. Structurally, nouns in French co-occur consistently with a preceding determiner. For example, English sentences, such as "food is expensive" and "mothers work a lot," must have a preceding determiner in French (*la nourriture est chère*; *les mères travaillent beaucoup*). These same sentences in German are comparable with those in English in that the nouns are not preceded by any determiner. In this sense, determiner–noun co-occurrence is more restrictive in French than in English and German. Verbs in French are frequently preceded by a subject pronoun. However, unlike nouns, verbs can occur without a preceding function word, such as imperatives (e.g., *Regarde*: "look") and yes–no questions involving verb–subject inversion (e.g., *veux (want)-tu(you) le lapin?*, "do you want the bunny"). Verbs can also follow various other kinds of words (e.g., nouns, adjectives, adverbs, and auxiliaries). In the speech of two Canadian French-speaking mothers to their 8-month-old infants (Cécyre & Shi, 2005), determiners were mostly followed by nouns or adjectives, with nouns being dominant (71%), and subject pronouns were mostly followed by verbs, auxiliaries, or other kinds of pronouns, with verbs being the most common (59%). In addition, one quarter of verbs occurred with no preceding word (i.e., utterance-initial or in isolation), whereas only about 2% of the nouns occurred utterance-initially or in isolation. Given our hypothesis that infants use the distributional patterns of function words with the adjacent nonfunction words for grammatical categorization, we expected noun categorization to be easier than verb categorization.

Third, we asked whether function words alone can allow infants to categorize the adjacent novel words. Recent analyses of infant-directed speech suggest that nouns and verbs are marked by differential prosodic cues in the input (Conwell & Morgan, 2009; Shi & Moisan, 2008). Prosodic differences associated with nouns and verbs in sentences were also reported in Nazzi, Dilley, Jusczyk, Shattuck-Hufnagel, and Jusczyk (2005). In previously published noun–verb categorization studies, the nonwords may have had differential prosodic characteristics when used as different categories, and such cues may have affected infants' performance. In this study, we balanced the prosodic properties of the nonword stimuli in noun versus verb uses so that the only cue to categorization was the preceding function word.

Finally, we used a different testing procedure (visual fixation procedure, Cooper & Aslin, 1990) than the Headturn Preference Procedure used by Höhle et al. (2004) and Mintz (2006) to see whether categorization results could be obtained across different testing settings.

METHODS

Participants

Thirty-two monolingual Canadian French-learning 14-month-old infants completed the study. These infants were divided into two familiarization groups (Det + N group: 10 boys, 6 girls; *M* age = 14 months 15 days; range = 14 months 2 days to 14 months 29 days; Pro + V group: 9 boys, 7 girls; *M* age = 14 months 14 days; range = 14 months 2 days to 15 months). Another 20 infants were tested but their data were excluded from the analysis due to fussiness (8), looking time too short (2 sec per trial or less for the initial seven trials; 6 infants), failure to stay on task (no recovery of looking time to the posttrial; 5 infants), and parental interference (1).

Stimuli

We created two nonsense words, *mige*, *crale*, which appeared as nouns with three determiners (i.e., Det + N: *des* [''some,'' indefinite determiner], *ton* [''your''], and *le* [''the'']), and as verbs with three subject pronouns (i.e., Pron + V: *je* [''I''], *il* [''he''], and *tu* [''you'']). Thus, there were a total of 12 two-word utterances. The function words are among the most frequent words in spoken French.[1] Furthermore, we chose them because determiners most often co-occur with nouns, and pronouns with verbs.

A native Canadian French female speaker recorded multiple repetitions of the 12 utterances in the infant-directed speech style in an acoustic chamber (sampling frequency 44.1 kHz, bit rate 16 bits). Utterances formed with two thirds of the determiners and the pronouns (*des*, *ton*, *je*, and *il*) were used for the familiarization phase, and the remaining one third of the determiners and pronouns (*le* and *tu)* were reserved for the test phase. The final stimuli consisted of three exemplars for each of the eight familiarization utterances (*ton mige*, *ton crale*, *des miges*, *des crales*; *je mige*, *je crale, il mige*, *il crale*), and six exemplars for each of the four test utterances (*le mige*, *le crale*; *tu miges*, *tu crales*). The stimuli were carefully chosen to ensure that

[1]Based on the Beauchemin adult-directed spoken corpus (Beauchemin, Martel, & Théoret, 1992; one million word tokens), the frequencies of the words are: *des* 8760, *ton* 1812, *le* 63979, *je* 30262, *il* 28144, and *tu* 10324. The average type-token ratio for all words is 1:8.83. Based on the speech of two Canadian French-speaking mothers to their 8-month-old infants (Cécyre & Shi, 2005) these words have the following frequencies: *des* 23, *ton* 15, *le* 51, *je* 25, *il* 21, and *tu* 75 (2,549 total word tokens). The average type-token ratio for all words is 1:5.58. Given that function words are a small set of closed-class words that form the structural frames of phrases and sentences, the token frequencies of these function words should increase drastically in larger infant-directed corpora.

TABLE 1
Average Acoustic Values (and Standard Deviations) Across Multiple Exemplars of the Nonword and Function Word Stimuli

Acoustic Measure	*Mean for Noun Uses*	*Mean for Verb Use*	*p (One-Way ANOVA)*
Total utterance duration (sec)	1.085 (.186)	1.089 (.191)	.932
Nonword (NW) duration (sec)	.849 (.153)	.867 (.138)	.662
Function word duration (sec)	.233 (.065)	.222 (.069)	.57
Function word mean pitch (Hz)	223.845 (25.342)	227.639 (20.753)	.573
Function word mean intensity (dB)	58.426 (14.559)	60.442 (10.132)	.270
NW vowel duration (sec)	.486 (.178)	.485 (.174)	.985
NW vowel mean pitch (Hz)	361.447 (116.744)	369.389 (102.812)	.804
NW vowel mean intensity (dB)	70.036 (5.236)	71.623 (4.168)	.251

the prosodic properties of the nonwords were comparable in noun and verb uses. Mean acoustic measures are shown in Table 1. All comparisons are nonsignificant, as indicated by the statistics in Table 1. We needed this control to test whether the preceding function words alone can drive successful categorization of nouns and verbs.

We used water bubble sounds as pre- and posttest materials. The pretest trial acquainted the infant with the procedure. The posttest trial allowed us to determine whether the infant was on task throughout testing. Looking time to the posttest trial should recover relative to the last test trial because the bubble sounds were distinct from the speech stimuli. A colorful checkerboard was presented during each trial. Auditory stimuli and the visual display were presented from the same physical location. They always began and ended simultaneously during each trial, a feature that was helpful for engaging the infants on the task and for inducing reliable looking responses. Finally, we used bird songs accompanied by a zooming star as the attention getter, to attract the infant's attention between trials.

Design

The experiment had two conditions (16 infants each). Infants in the Det + N familiarization condition were familiarized with trials presenting the exemplars of the following utterances: *ton mige*, *ton crale*, *des miges*, and *des crales*.[2] Infants in the Pron + V familiarization condition were familiarized with trials presenting the exemplars of the following: *je mige*, *je crale*, *il mige*, and *il crale*. For instance, a Det + N familiarization trial would randomly draw exemplars of the four utterances containing *ton* and *des*. To ensure

[2]Note that the "s" following the nonwords is not pronounced in French.

stimuli variability, the randomization within each trial was conducted in blocks such that each block contained one exemplar of each of the four utterances.

The test phase immediately followed the familiarization phase. Infants from both familiarization conditions were tested with the same stimuli. There were two types of test trials: Same Category versus Different Category. The Same Category trials involved the use of the familiarization nonwords in new function word context that respected the syntactic category of the familiarization phase. In the Different Category trials, the nonwords appeared with a function word that violated the syntactic category use of the familiarization phase. Specifically, for the infants in the Det + N familiarization condition, a test trial presenting *le mige* and *le crale* exemplars would be the Same Category type, and a trial presenting *tu miges* and *tu crales* exemplars would be the Different Category type. The reverse was the case for the Pron + V familiarization condition, with *tu miges* and *tu crales* as the Same Category type, and *le mige* and *le crale* the Different Category type. The test phase began with a Same Category or Different Category trial, counterbalanced across infants. Same Category trials and Different Category trials were presented in an alternating fashion for a total of 10 trials. The interval between any two utterances within a trial was 700 msec.

Procedure

Infants were individually tested in a visual fixation procedure (Cooper & Aslin, 1990). The infant sat on the parent's lap in a chamber approximately 2 m from a monitor and loudspeakers (placed at the same location). The parent, who was asked not to interact with the infant, heard masking music through noise-cancellation headphones. The researcher, blinded to the order of stimuli presentation, observed the infant in an adjacent room through a closed-circuit TV, and administered the experiment using a testing program (Cohen, Atkinson, & Chaput, 2000). She pressed down a computer key whenever a look toward the monitor occurred. The software presented trials containing the audio-visual stimuli, recorded all looking times online, and calculated automatically the total looking time across familiarization trials. Once the looking time reached the predetermined familiarization criterion (30 sec), the software moved the experiment automatically to the test phase. Each trial was initiated contingent upon the infant's look toward the monitor. A trial would terminate when the infant looked away for at least 2 sec. If the infant kept looking at the monitor or looked away for less than 2 sec during a trial, the trial would continue until the maximum trial length of 19.8 sec was reached.

Predictions

If infants can categorize individual determiners as one common class, then after being familiarized to *des* and *ton* phrases, they should show differential looking responses to the test trial type containing the new determiner *le* versus that presenting the pronoun *tu*. Similarly, the ability to perceive different pronouns as a single category would be indicated by differential looking times to the test trial types containing *le* versus *tu* in the reverse direction after the pronoun familiarization phase with *je* and *il* utterances.

The design of our experiment can also provide information concerning whether infants are able to use function words for categorizing the subsequent novel words to rudimentary noun and verb classes. If infants could discriminate the two types of test trials (i.e., Same Category versus Different Category), we could interpret the result as supporting noun–verb categorization. Importantly, both test trial types included the familiarization nonwords occurring with new function words not heard during familiarization. Therefore, differential responses could not be due to pure memory of the familiarization materials. But rather, they may indicate infants' expectation about what other function words could co-occur with the familiarized nonwords, i.e., the categories of the nonwords. We predicted that the ability to categorize novel words to nouns would be stronger than that to verbs, based on the linguistic characteristic that determiners and nouns co-occur more consistently than do pronouns and verbs in French. Infants in the Det + N familiarization condition should show a significant response difference for the two test trial types. Infants in the Pron + V familiarization condition were expected to show less or no difference.

RESULTS AND DISCUSSION

Each infant's average looking times for the Same Category trials and for the Different Category trials during the test phase were calculated. The data for all infants were analyzed in a 2 × 2 analysis of variance (ANOVA), with Test Category (Same Category versus Different Category) as the within-subject factor, and Familiarization (Det + N versus Pron + V) as the between-subject factor. The results showed no effect of Familiarization, $F(1, 30) = 1.053$, $p = .313$, and a near significant effect of Test Category, $F(1, 30) = 4.087$, $p = .052$. However, the interaction between Test Category and Familiarization approached significance, $F(1, 30) = 3.760$, $p = .062$. Follow-up paired *t*-tests assessed the performance of each familiarization group. For the Det + N familiarization group, looking time was significantly longer for the Different Category trials ($M = 7.671$ sec, $SE = 1.107$ sec)

than for the Same Category trials (M = 5.393 sec, SE = .717 sec), $t(15) = -2.593$, p = .02. In contrast, the Pron + V familiarization group showed no difference for the Same Category (M = 7.626 sec, SE = .747 sec) versus Different Category trials (M = 7.674 sec, SE = .86 sec), $t(15) = -.064$, p = .95. These results as shown in Figure 1 agree with our predictions.

We conducted the same analyses with the first test trial of each type removed, as is often done in this procedure (e.g., Cooper & Aslin, 1994; Marquis & Shi, 2008). The response pattern was comparable, and the results more robust. There was again no main effect of Familiarization, $F(1, 30) = 2.176$, p = .151. The main effect of Test Category did not reach significance, $F(1, 30) = 3.112$, p = .088. However, the interaction of these two main factors was significant, $F(1, 30) = 6.245$, p = .018. Follow-up t-tests for the Det + N familiarization group yielded a significantly longer looking time for the Different Category trials (M = 6.988 sec, SE = 1.212 sec) than for the Same Category trials (M = 4.405 sec, SE = .561 sec), $t(15) = -2.759$, p = .015. The two test trial types did not differ for the Pron + V familiarization group (Same Category: M = 7.714 sec, SE = .911 sec; Different Category: M = 7.269 sec, SE = 1.040 sec), $t(15)$ = .579, p = .571. These results further confirmed our predictions.

Our results show that the Det + N familiarization infants treated the new determiner in the Same Category trials during the test phase as the same kind of words as the determiners in the familiarization phase. That is, infants generalized different determiners to one common class, but considered pronouns as not belonging to that class. Note that the determiners were produced with equivalent prosody as the pronouns (see Table 1). Thus,

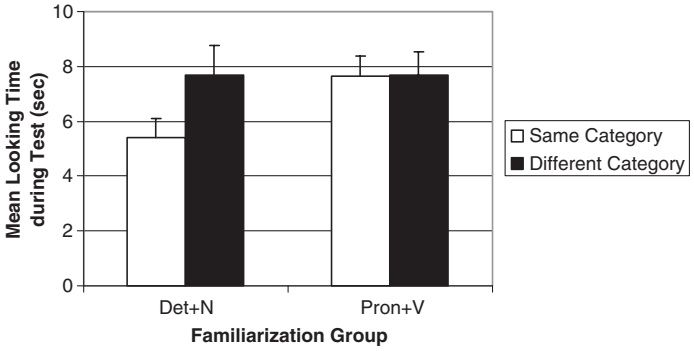

Figure 1 Mean looking times (and standard errors) during the Same Category test trials and the Different Category test trials for the Det + N familiarization condition (left columns) and the Pron + V familiarization Condition (right columns).

infants had specific representations for individual function words, and seemed to expect the determiner class to co-occur with the same nonwords. They appeared surprised by the use of the familiarized nonwords in conjunction with a function word that is not a determiner (here, a pronoun), suggesting that infants might have used determiners to categorize novel words to nouns.

In contrast, the Pron + V familiarization infants did not generalize the pronouns in the familiarization and test phases as a common class. Not only was there no response difference for the two test trial types after familiarization, infants also looked long toward both the Same Category and Different Category trials. Independent *t*-tests showed that looking times to the Same Category trials and the Different Category trials in the Pron + V familiarization condition were significantly longer than to the Same Category trials in the Det + N familiarization condition, $t(30) = -3.093$, $p = .004$ and $t(30) = -2.424$, $p = .022$, respectively, but not different from the Different Category trials in the Det + N familiarization condition, $t(30) = -.479$, $p = .635$ and $t(30) = -.176$, $p = .861$, respectively. Thus, Pron + V familiarization infants treated both the determiner and the pronoun in the test phase as equally new items unrelated to the pronouns heard during familiarization. The overall long looking to both test trial types suggests that infants can represent distinctly the individual function words in that condition. However, the lack of looking difference to the two test trial types means that infants failed to generalize the pronouns into one class and may have also failed to categorize novel words to verbs.

Our study goes beyond existing research on infant grammatical categorization in that we showed that as young as 14 months of age, infants already have the knowledge of some refined functional categories, such as the determiner class. It is interesting to note that children's early speech typically lacks function words (e.g., Bassano, Eme, & Champaud, 2005; Brown & Fraser, 1963), and that determiners appear in production many months later than the age that we examined in this study. Our results demonstrate that categorization of function words begins much earlier than previously understood. This has implications for acquisition theories. For instance, knowledge of specific functional categories may guide infants' acquisition of other syntactic categories, for example verb meanings, as postulated in the model of syntactic bootstrapping (Gleitman, 1990). While semantically based theories of category learning place content words as those to be first assigned syntactic categories, we showed that infants' initial categories include functional categories. This ability represents a logical developmental progression during and shortly after the first year of life, in that infants can already make the broad form-based categorization of lexical versus function words at birth (Shi et al., 1999), and that they begin to recognize, represent and store

individual function words after 6 months of age (e.g., Hallé et al., 2008; Höhle & Weissenborn, 2003; Shady, 1996; Shafer et al., 1998; Shi, Cutler et al., 2006; Shi & Lepage, 2008; Shi, Marquis et al., 2006; Shi, Werker et al., 2006). It is reasonable that formal categorical learning involving the relation between functional categories and lexical categories occurs subsequently. The present study shows that knowledge about rudimentary functional categories is indeed present at the onset of syntactic acquisition.

Our study also extends the previous work on grammatical categorization to French-learning infants. Unlike previous studies, we carefully balanced the acoustic properties of the novel words in noun versus verb uses. Hence, we showed that function words alone could guide basic grammatical categorization, and that this effect is probably related to the consistent co-occurrence of function words with the neighboring words in the natural input environment. More specifically, the results showed that the bigram cue involving a preceding function word may be sufficient for initial noun categorization in French but not for verb categorization. This suggests that infants would need to seek other cues for categorizing verbs, e.g., trigrams (such as frequent frames), phonological/acoustic cues of nouns versus verbs, or combinations of different cues. Indeed, corpus studies of parental speech have shown that frequent frames involving both the preceding and following contextual words (mostly function words) provide more restrictive contexts and can yield better categorization performance for both nouns and verbs than do bigrams (Chemla, Mintz, Bernal, & Christophe, 2009; Mintz, 2003). The learning system may be opportunist, in that when one cue is weaker than another, more reliance may be on the stronger cue. In connectionist studies, the integration of different cues was shown to be efficient for the learning system to form basic grammatical categories (e.g., Monaghan et al., 2007; Shi et al., 1998). More empirical studies with infants are needed to ascertain how exactly the categorization mechanism works under different input conditions.

Our results can be considered within the larger literature on noun versus verb acquisition. Many studies in various languages showed that nouns appear earlier than verbs in children's vocabulary (e.g., Bassano et al., 2005; Gentner, 1982; Goldin-Meadow, Seligman, & Gelman, 1976). This developmental difference was explained in conceptual terms (Gentner, 1982), in that nouns have more direct word-to-world references, whereas verbs denote more complicated predicative concepts. This idea was recently confirmed in a study of novel word mapping and extension with English-, Japanese- and Chinese-learning children (Imai et al., 2008). Nouns also seem to be at an advantage at other levels of processing. For example, segmentation of nouns begins from 6 months of age, based on results primarily from English (e.g., Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Jusczyk & Aslin, 1995;

Jusczyk, Houston, & Newsome, 1999). Verb segmentation appears to be delayed for at least a few months relative to nouns in English and French (Marquis & Shi, 2008; Nazzi et al., 2005). Segmentation may also be affected by the degree of morphological complexity for nouns versus verbs. More work with children learning other languages is needed to determine whether nouns are at an advantage across different language acquisition tasks including word segmentation and grammatical categorization.

It should be noted that the distributional patterns related to nouns and verbs in French contain some degree of noise. One potential problem concerning the usefulness of determiners is that the French definite articles *le/la/les* can also serve as object pronouns before verbs (e.g., *Je le vois*, "I see him"), which creates ambiguity with determiner use before nouns (e.g., *le chien*, "the dog"). However, the two kinds of uses are not balanced in actual speech. The count of *le* (in footnote 1), which we took from the adult spoken database (Beauchemin et al., 1992), is only the determiner use. If we consider the determiner and object pronoun uses together, the determiner use is 87%, whereas the object pronoun use is 13%. We also found a similar pattern in the available infant-directed Canadian French corpus (Cécyre & Shi, 2005): The use of *le/la/les* as determiners is 89% and as object pronouns 11%. Thus, the determiner use is largely dominant in comparison with the ambiguous object pronoun use. Given the existing evidence from artificial language experiments that infants can learn dominant input patterns in the presence of a certain degree of noise (Gomez & Lakusta, 2004), it is plausible that the largely consistent use of *le/la/les* as determiners can guide infants' initial rough categorization of nouns despite certain inconsistent, ambiguous cases.

One interesting structural characteristic of French is that there are both pre- and postnominal adjectives. Prenominal adjectives can intervene between the determiner and the noun. Such occurrences, if dominant, could pose a challenge to the use of determiners for categorization of nouns. However, based on our analysis of the infant-directed corpus (Cécyre & Shi, 2005), the proportion of tokens following determiners which were adjectives is 26%, and these belong to a small set of adjectives (e.g., *petit* and *gros*). In contrast, 71% of word tokens following a determiner were nouns, which is dominant relative to the adjective occurrences.

An important issue concerns the nature of infants' knowledge shown in this study. Some child language researchers maintain that infants' initial knowledge of language is not syntactic, and that abstract grammatical categories are established much later, about 2–3 years of age (e.g., Pine & Lieven, 1997; Tomasello, 2000). Other researchers claim that category learning is about mapping the input to preexisting abstract syntactic categories (e.g., Valian, 2009). Whereas the present study is not a test of the existence

of innate syntactic categories, it does show that infants perform formal, abstract analyses of the input (their distributional analyses concern relations between function words and lexical words), and that they can generalize the relational patterns to rudimentary word classes and extend them to novel words. This process resembles the way syntactic categories per se are defined in the mature grammar. In this sense, our results are compatible with the view of continuity. Nevertheless, truly abstract grammatical categories imply the knowledge of the full range of syntactic category relations, which infants have yet to develop at a later age.

In sum, this study aimed at better understanding early grammatical categorization. We showed that 14-month-olds not only perceive and represent specific function word forms, but they also have knowledge about function words that are syntactic in nature. In particular, we demonstrated that infants at this age already have the rudimentary syntactic category of determiners. Furthermore, our results suggest that infants appear to attend to the distribution between frequent function words and the following words. The largely consistent distribution of determiners with adjacent nouns in the input seems sufficient for French-learning infants' categorization of novel words to nouns. Overall, our findings support the general hypothesis (e.g., Christophe, Guasti, Nespor, Dupoux, & Ooyen, 1997; Shi et al., 1998) that function words/morphemes are important for bootstrapping early language development.

ACKNOWLEDGMENTS

We thank all parents and infants who participated in this study. The results of this study were presented at the XVIth Biennial International Conference on Infant Studies, Vancouver. We also thank the organizers of the workshop ''Current Issues in Language Acquisition: Artificial and Statistical Language Learning'' for inviting us to discuss our studies. This research was supported by funding from SSHRC, NSERC, and CFI.

REFERENCES

Bassano, D., Eme, P. E., & Champaud, C. (2005). A naturalistic study of early lexical development: General processes and inter-individual variations in French children. *First Language*, *25*, 64–101.

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.

Bernal, S., Lidz, J., Millotte, S., & Christophe, A. (2007). Syntax constrains the acquisition of verb meaning. *Language Learning & Development*, *3*, 325–341.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech stream segmentation. *Psychological Science*, *16*, 298–304.

Braine, M. (1976). Children's first word combinations. *Monographs of the Society for Research in Child Development*, *41* (1, Serial No. 164).

Brown, R., & Fraser, C. (1963). The acquisition of syntax. In C. Cofer & B. Musgrave (Eds.), *Verbal behaviour and learning: Problems and processes* (pp. 158–196). New York: McGraw-Hill.

Cécyre, M.-J., & Shi, R. (2005). *Speech rate in maternal speech to French-learning infants*. Paper presented at the 15th meeting of the Canadian Society for Brain, Behaviour and Cognitive Science, Montreal, Canada.

Chemla, E., Mintz, T., Bernal, S., & Christophe, A. (2009). Categorizing words using 'Frequent Frames': What cross-linguistic analyses reveal about distributional acquisition strategies. *Developmental Science*, *12*, 396–406.

Christophe, A., Guasti, T., Nespor, M., Dupoux, E., & Ooyen, B. V. (1997). Reflections on phonological bootstrapping: Its role for lexical and syntactic acquisition. *Language and Cognitive Processes*, *12*, 585–612.

Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2000). *Habit 2000: A new program for testing infant perception and cognition*. (Version 2.2.5c) [Computer software]. Austin, TX: The University of Texas.

Conwell, E., & Morgan, J. L. (2009). Is it a noun or is it a verb? Resolving the ambicategoricality problem. Manuscript submitted for publication.

Cooper, R. P., & Aslin, R. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*, 1584–1595.

Cooper, R. P., & Aslin, R. N. (1994). Developmental differences in infant attention to the spectral properties of infant-directed speech. *Child Development*, *65*, 1663–1677.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kuczaj II (Ed.), *Language development, Volume 2: Language, thought and culture* (pp. 301–334). Hillsdale, NJ: Lawrence Erlbaum.

Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language*, *32*, 249–268.

Gleitman, L. R. (1990). The structural sources of verb meaning. *Language acquisition*, *1*, 3–55.

Goldin-Meadow, S., Seligman, M., & Gelman, R. (1976). Language in the two-year-old. *Cognition*, *4*, 189–202.

Gomez, R., & Lakusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science*, *7*, 567–580.

Hallé, P., Durand, C., & de Boysson-Bardies, B. (2008). Do 11-month-old French infants process articles? *Language and Speech*, *51*, 23–44.

Höhle, B., & Weissenborn, J. (2000). The origins of syntactic knowledge: Recognition of determiners in one year old German children. In *Proceedings of the 24th Annual Boston Conference on Language Development*. Somerville, MA: Cascadilla Press.

Höhle, B., & Weissenborn, J. (2003). German learning infants' ability to detect unstressed closed class elements in continuous speech. *Developmental Science*, *6*, 122–127.

Höhle, B., Weissenborn, J., Kiefer, D., Schulz, A., & Schmitz, M. (2004). Functional elements in infants' speech processing: The role of determiners in segmentation and categorization of lexical elements. *Infancy*, *5*, 341–353.

Imai, M., Li, L., Haryu, E., Okada, H., Hirsh-Pasek, K., Golinkoff, R. M. & Shigematsu, J. (2008). Novel noun and verb learning in Chinese-, English-, and Japanese-speaking children. *Child Development*, *4*, 979–1000.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*, 1–23.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, *39*, 159–207.

Maratsos, M. P., & Chalkley, M. A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K. Nelson (Ed.), *Children's Language* (Vol. 2, pp. 127–214). New York: Gardner Press.

Marquis, A., & Shi, R. (2008). Segmentation of verb forms in preverbal infants. *Journal of the Acoustical Society of America*, *123*, EL105–EL110.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, *90*, 91–117.

Mintz, T. H. (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* (pp. 31–63). New York: Oxford University Press.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition*, *96*, 143–182.

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology*, *55*, 259–305.

Nazzi, T., Dilley, L. C., Jusczyk, A. M., Shattuck-Hufnagel, S., & Jusczyk, P. W. (2005). English-learning infants' segmentation of verbs from fluent speech. *Language & Speech*, *48*, 279–298.

Pine, J., & Lieven, E. (1997). Slot and frame patterns in the development of the determiner category. *Applied Psycholinguistics*, *18*, 123–138.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, *22*, 425–469.

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old children. *Science*, *274*, 1926–1928.

Shady, M. (1996). *Infants' sensitivity to function morphemes*. Doctoral dissertation, University at Buffalo, The State University of New York, Buffalo, NY.

Shafer, V., Shucard, D., Shucard, J., & Gerken, L. (1998). An electrophysiological study of infants' sensitivity to the sound patterns of English speech. *Journal of Speech, Language, and Hearing Research*, *41*, 874–886.

Shi, R., Cutler, A., Werker, J., & Cruickshank, M. (2006). Frequency and form as determinants of functor sensitivity in English-acquiring infants. *Journal of the Acoustical Society of America*, *119*, EL61–EL67.

Shi, R., & Lepage, M. (2008). The effect of functional morphemes on word segmentation in preverbal infants. *Developmental Science*, *11*, 407–413.

Shi, R., Marquis, A., & Gauthier, B. (2006). Segmentation and representation of function words in preverbal French-learning infants. In D. Bamman, T. Magnitskaia, & C. Zaller (Eds.), *BUCLD 30: Proceedings of the 30th annual Boston University conference on language development* (Vol. 2, pp. 549–560). Boston, MA: Cascadilla Press.

Shi, R., & Moisan, A. (2008). Prosodic cues to noun and verb categories in infant-directed speech. In *BUCLD 32: Proceedings of the 32th annual Boston University conference on language development*. Boston, MA: Cascadilla Press.

Shi, R., Morgan, J., & Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *Journal of Child Language*, *25*, 169–201.

Shi, R., Werker, J., & Cutler, A. (2006). Recognition and representation of function words in English-learning infants. *Infancy*, *10*, 187–198.
Shi, R., Werker, J., & Morgan, J. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, *72*, B11–B21.
Tomasello, M. (2000). The item based nature of children's early syntactic development. *Trends in Cognitive Sciences*, *4*, 156–163.
Valian, V. (2009). Innateness and learnability. In E. Bavin (Ed.), *Handbook of child language* (pp. 15–34). Cambridge: Cambridge University Press.
Valian, V., & Coulson, S. (1988). Anchor points in language learning: The role of marker frequency. *Journal of Memory and Language*, *27*, 71–86.

[This paper was published in Infancy, 15(5), 517-533]