Grammatical Gender Categorization in Infants

Marilyn Cyr and Rushen Shi
Université du Québec à Montréal

1. Introduction

Learning to assign words to grammatical categories is an important task during language acquisition. Phrases and sentences are not defined by specific words such as *dog* or *walk*, but rather by categories such as nouns, verbs, prepositions and determiners. This aspect is tied to the productive nature of human languages. All syntactic theories, regardless of their differences, accept the notion of grammatical categories as the basic constructs. Thus, with a lexicon organized into grammatical categories, a small set of syntactic structures can generate a nearly infinite number of sentences. This knowledge is present in all native speakers of a natural language, allowing them to produce and comprehend sentences that they have never heard. Grammatical category learning would have been trivial if learners were explicitly taught category labels as is usually the case during second language teaching. Children, however, do not receive explicit instructions on grammatical categories. They must discover the category membership from the speech input. How they achieve this task has been the focus of many acquisition researchers.

Different theoretical models have been proposed in the literature for the understanding of grammatical category learning. One type of models focuses on the role of semantics (e.g., Pinker, 1984; Gentner, 1982). Infants may rely on meanings to acquire the grammatical categories of their first words. For example, objects can be nouns, action can be verbs, and properties can be adjectives. They may use meanings to map words to innate abstract syntactic categories (Pinker, 1984), or to create grammatical categories (Gentner, 1982).

Distributionally based theories are an alternative type of approach to the issue of grammatical categorization (e.g. Maratsos & Chalkley, 1980; Braine, 1992). According to this theory, rather than beginning to form syntactic categories by using referential information, infants first rely on distributional information (e.g., co-occurrence patterns, the relative position of the words in a sentence) in the speech input. At a later stage of learning, correlation between distributional and referential information can contribute further syntactic acquisition.

Another theoretical approach to grammatical category learning works with the idea that there is a link between the syntactic structure and the phonological properties in natural languages. With regards to grammatical categories, it was proposed that input speech to children contains distinct phonological and/or prosodic characteristics for basic categories (e.g., Christophe, Guasti, Nespor, Dupoux & Ooyen, 1997; Morgan, 1986; Shi, Morgan, & Alloppenna, 1998).
These theoretical models are all concerned with the question of how infants break into the system of grammatical categories. Therefore, it is necessary to test the models by examining infants at the earliest stage of category learning. Experimental studies with infants are relatively sparse. The youngest age that has been tested was 1-3-day-old newborns (Shi, Werker, & Morgan, 1999), who showed evidence of categorically discriminating content words versus function words based on prosodic and phonological cues. This early ability is reasonable given that the distinction between content words and function words/morphemes constitutes the most fundamental, binary category distinction in all languages, which are supported by universal cues (Shi, Morgan, & Allopenna, 1998).

Categorization of more refined categories such as nouns versus verbs in infants has been examined in recent experiments. It was shown that German-learning infants can categorize pseudo-words into nouns when those are preceded by a determiner. Using a Headturn preference procedure, Höhle, Weissenborn, Kiefer, Schulz, and Schmitz (2004) familiarized infants with pseudo-words preceded either by a determiner or a pronoun in German. The test phase consisted of sentences with the pseudo-words used in both noun and verb contexts. Infants by the age of 14-16 months showed a difference in listening time between the two test trial types when they were familiarized in the noun condition, but did not show such a difference when they were familiarized with the verb context. The authors concluded that infants categorized the nouns based on the co-occurrence distributional pattern between determiner and nouns in German. However, since other uncontrolled cues (e.g. possible prosodic differences when the pseudo-words were produced as nouns versus as verbs) may have also affected their results, it is unclear whether their categorization results was purely based on distributional information.

Another study by Mintz (2006) showed evidence of verb categorization in 12-month-old English-learning infants. In that study, infants were familiarized with sentences containing pseudo-words that occurred either in noun frequent frames (e.g. the_in) or in verb frequent frames (e.g. to_it). During the test phase, the same pseudo-words were presented in novel sentences: grammatical (i.e. the pseudo-words occurred in a new frequent frame supporting the category that was consistent with the familiarization) versus ungrammatical (i.e. the pseudo-words occurring in a frequent frame that supported a category inconsistent with the familiarization). The infants only showed a differential looking patterns for the pseudo-words familiarized in verb frames. The author suggested that the results were predicted by the frequent frames (distributional cues) associated with verbs in English. However, like the Höhle et al. (2004) study, the prosody of the pseudo-words in noun versus verb frames was not controlled. Therefore, in both Höhle, et al. (2004)’s and Mintz (2006)’s studies, it is unclear whether categorization of pseudo-words was based entirely on distributional information, or whether prosodic cues in the stimuli may have affected the results.

There are other studies that manipulated more specifically the phonological and distributional cues of the stimuli (Gerken, Wilson & Lewis, 2005; Gomez & Lakusta, 2004). In an artificial language experiment, Gomez and Lakusta (2004)
trained infants with an aX bY paradigm, where a and b words were comparable to function words, while X and Y words were like content words. They showed that 12-month-olds can form categories by associating distributional information of a and b words with phonological features of X and Y words, and that they can generalize this association to novel instances.

In another study, Gerken et al. (2005) trained 17-month-old American infants with the Russian grammatical gender paradigm. They found that infants needed two cues to form categories, i.e., the contextual distribution of the functional morpheme (gender suffix), and a phonological cue at the end of the noun stem. When only the distributional information was present (i.e., the suffix co-occurring with stem noun), without the phonological cue on the stem, infants failed to show grammatical gender categorization. This experiment showed that infants required the association of function words/morphemes with the phonological cues on the content words/morphemes. That is, grammatical categorization based on distributional information alone may be insufficient. Learning might require the collaboration of prosodic support.

Syntactically, grammatical categories are defined in terms of structural relations, which are distributional in nature. For example, the functional categories such as determiners and pronouns are defined in terms of their distributional relations with content-word categories such as nouns and verbs. Structural relations exist with or without phonological or prosodic correlates. Infants must ultimately learn to categorize known and novel words distributionally, without requiring phonological/acoustic cues on content words. In this sense, the distributionally based categorization represents a rudimentary kind of syntactically or structurally based knowledge. It may reflect a higher-level knowledge than one that requires an association, such as the associations shown in Gomez and Lakusta (2004) and Gerken et al. (2005). From the learning point of view, if infants can perform distributionally based categorization from the earliest stage, category learning would be more direct to the core of syntax than for mechanisms that rely on other types of information such as semantics.

The aim of the present study is therefore to determine when the ability of categorization based on distributional information alone (i.e., structural relations) emerges in infancy.

Grammatical genders in French can serve as a good testing case for this research question. In French, nouns are subcategorized in masculine and feminine and the determiners agree with masculine and feminine nouns. A feminine noun is preceded by a feminine determiner (e.g., une maison, la maison) and a masculine noun is preceded by a masculine determiner (e.g., un ballon, le ballon). There are many masculine and feminine nouns that show some phonological properties that associate with gender. For example, the analyses of the noun corpus of a French dictionary revealed that more than 60% of the nouns carry an ending that associates in a systematic way with gender (Tucker, Lambert, and Rigault, 1977; in Seigneuric, Zagar, Meunier & Spinelli, 2007), such as -ette ending (e.g., pousette) that co-occurs with feminine, and -in
ending (e.g. *bain*) that co-occurs with masculine gender. However, such associations are far from perfect. For example, *squelette* is a masculine noun in French, but it carries an ending that is highly frequent in feminine nouns. Furthermore, many nouns carry a neutral ending, and neutral endings do not co-occur in a systematic manner with a specific gender. Therefore, although infants may perform categorization by associating determiner distributions with the phonological gender cues in some nouns, they still must learn to categorize without relying on the phonological cues.

In this study we ask whether French-speaking infants are able to categorize novel nouns as masculine or feminine based exclusively on contextual distributional cues. More precisely, we ask if infants are able to categorize novel nouns according to the preceding determiner (feminine or masculine) and generalize to other determiners of the same gender. A recent categorization study showed that French-learning 14-month-olds can group determiners as one grammatical category, and treat other function words like pronouns as distinct (Shi and Melançon, in press). It is not yet known whether infants at this age subcategorize determiners into gender classes.

2. Experiment 1
2.1. Participants and Stimuli

Participants were eleven 14-month-old (Mean age = 444.36 days; SD = 5.94; range: 432-453) monolingual Quebec-French-learning infants. The stimuli for this experiment were French determiners and pseudo-nouns. Pseudo-nouns were used for assessing infants’ capacity to categorize grammatical genders to novel words. They also allowed us to better control for the word forms in the goal to neutralize any phonological cues to gender. Furthermore, using pseudo-words allowed controlling for prior lexical knowledge of the children.

To ensure that the pseudo-words we chose for this experiment did not contain phonological cues to grammatical gender, a pre-experiment with French-speaking adults was first carried out to determine pseudo-words that were equally possible in both genders. This allowed us to eliminate the potential effect of phonological cues on the gender of the nouns and test the distributional contribution alone. A series of pseudo-words were created and verbally presented to 10 participants. The participants were asked to verbally create a sentence using each pseudo-word as a noun. They were then asked to judge if each pseudo-word was noun-like in French, and if it was a typical-sounding noun. The implicit objective of this experiment was to induce gender production. In particular, if a pseudo-word is phonologically biased toward one of the two genders, consistency towards the use of a determiner of one gender should be high between subjects. However, if a pseudo-word is equally possible in both genders, the consistency between participants should be low, i.e., the categorization of the pseudo-word with masculine versus feminine determiners should be at the chance level. From this pre-experiment, four pseudo-nouns that were judged as typical nouns in French and that had an ambiguous gender use
across adults (i.e., with comparable use of masculine or feminine determiners) were selected: mouveil, ravol, cagère, gombal.

A monolingual Quebec-French female speaker recorded the stimuli, using the infant-directed speech style. The recording was conducted in a sound proof recording booth, using a digital audio recorder Sound Device 702T at 48kHz sampling frequency and 24bits bit rate. The speech was then transferred digital-to-digital to the computer, and edited with Praat (version 4.5.26) (Boersma & Weenink, 2007). The four nonsense words were each produced multiple times in isolation and in noun phrases (preceded by four French determiners: two masculine determiners, un and le, and two feminine determiners, une and la).

Final stimuli set for the familiarization phase included three tokens of each of the following utterances: un mouveil, une mouveil, un gombal, une gombal, un ravol, une ravol, un cagère and une cagère. Stimuli for the test phase included three tokens of each of the following utterances: le mouveil, la mouvaille, le gombal, la gombale, le ravol, la ravole, le cagère and la cagère.

To avoid possible differential acoustic/prosodic cues in the pseudo-nouns when produced with masculine versus feminine determiners, we created all the stimuli by cross-splicing the pseudo-nouns that were produced in isolation with determiners of both genders. The same token of a pseudo-noun (gender neutral citation production) was conjoined with both masculine and feminine determiners which were spliced from noun phrases containing the corresponding pseudo-noun. Therefore, with these two controls (i.e., the phonological control and the prosodic/acoustic control), the only cue to gender present in the stimuli was the gender of the determiners (representing contextual distributional cue), i.e., there was no association between the determiners and any phonological or acoustic cues on the pseudo-nouns.

The visual stimuli consisted of a cartoon-like puppet standing in the center of the screen. The mouth movement of the puppet synchronized with the auditory stimuli. While “speaking”, the puppet also moved her hands, head and body. The visual stimuli were created using Adobe® Flash® CS3 professional software. We used a 30 fps frame rate and a render of 1360 x 768 pixels, which corresponds to the size of the screen on which the visual stimuli were presented. This animation was created to keep infants’ interested in the task.

2.2. Design and Predictions

The experiment included a familiarization phase and a test phase. During the familiarization phase the four pseudo-nouns were each preceded by a French indefinite determiner. Two of the pseudo-nouns were consistently preceded by the masculine determiner un, while the other two pseudo-nouns were preceded by the feminine determiner une. During the test phase the same pseudo-nouns were preceded by a French definite determiner (i.e., le – masculine, la – feminine). The test phase consisted of two test trial types: grammatical versus ungrammatical. In the grammatical test trials, two pseudo-nouns from the familiarization phase were each paired with a definite determiner, the gender of
which was consistent with the gender pairing during familiarization. For example, one pseudo-noun that was paired with a masculine indefinite determiner during the familiarization phase was paired with a masculine definite determiner in the grammatical trials during the test phase, and a second pseudo-noun that was paired with a feminine indefinite determiner during the familiarization phase was paired with a feminine definite determiner in the grammatical trials during the test phase. The ungrammatical test trials consisted of the other two pseudo-nouns, each of which was paired with an “incorrect” definite determiner, i.e., the determiner gender for these two nouns was inconsistent between the familiarization and test phases: Specifically, one of these two pseudo-nouns that was paired with a masculine indefinite determiner during the familiarization phase was then paired with a feminine definite determiner in the ungrammatical trials of the test phase, and one pseudo-noun that was paired with a feminine indefinite determiner during the familiarization phase was paired with a masculine definite determiner in the ungrammatical trials of the test phase. The determiner-noun pairings were counterbalanced across infants. Table 1 shows the pairing of one of the experimental conditions.

<table>
<thead>
<tr>
<th>Familiarization</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>un mouveil</td>
<td>Grammatical</td>
</tr>
<tr>
<td>un ravol</td>
<td>trials</td>
</tr>
<tr>
<td>une gombale</td>
<td>la gombale</td>
</tr>
<tr>
<td>une cagère</td>
<td>le cagère</td>
</tr>
<tr>
<td>Test</td>
<td>Ungrammatical</td>
</tr>
<tr>
<td>trials</td>
<td>trials</td>
</tr>
<tr>
<td>le mouveil</td>
<td>la ravole</td>
</tr>
</tbody>
</table>

Table 1. Familiarization and test materials for one condition.

The familiarization trials consisted of the presentation of the three tokens of each pseudo-noun paired with an indefinite determiner. The tokens had been randomly concatenated into trials of 15.1 seconds. When infants accumulated 90 seconds of looking, the test phase began. The test phase consisted of a total of 10 trials of a maximal length of 15.2 seconds. The two test trial types were presented in alternation. The first trial of the test phase was either grammatical or ungrammatical, counterbalanced across infants.

2.3. Procedure

Participants were tested individually in a preferential looking paradigm. In a sound-attenuated test booth the child sat on the parent’s lap. The parent listened
to masking music delivered by noise cancellation headphones. At approximately two meters in front of the child, there was a 42-inch LG 1360 x 768 TV monitor. The experimenter, who was blind to the audio-visual stimuli, observed the child’s eye movements in an adjacent room through a closed-circuit TV and pressed down a computer key whenever an eye fixation occurred. The experimental software (Cohen, Atkinson & Chaput, 2000) presented the stimuli and recorded all looking times. The experiment was infant-controlled. Each trial was initiated by the child, i.e., the experimenter waited for the child to look toward the monitor before launching a trial. Once a trial started, it continued until the infant stopped looking at the monitor for at least 2 seconds or until the maximum trial length was reached. Between trials an attention getter was played to attract the infant’s attention. The attention getter was an animation of a bird zooming in and out, synchronized with a whistle sound.

2.4. Results

A paired t-test was conducted to compare infants’ looking times while listening to the grammatical and the ungrammatical test trials. Infants’ looking time to the grammatical trials ($M=39.11$ sec; $SE=5.68$ sec) was similar to their looking time to the ungrammatical trials ($M=38.84$ sec; $SE=5.77$ sec), $t(10)=0.66$, $p=.948$, 2-tailed. Therefore, infants showed no evidence of gender categorization on the sole basis of determiner distribution with content words.

![Fig. 1. 14-month-old infants’ looking (listening) times (means and standard errors) to the two test trial types, grammatical (the determiner gender pairing consistent with the familiarization) versus ungrammatical (the determiner gender pairing inconsistent with the familiarization).](image)

3. Experiment 2

The 14-month-old infants in Experiment 1 showed no evidence of gender categorization. The goal of Experiment 2 was to examine whether slightly older infants can perform purely distributionally based gender categorization.
3.1. Participants, Stimuli, Design, Procedure and Predictions

Fourteen 17-month-old (Mean age = 532.5 days; SD = 14.75; range: 486-548) monolingual Quebec-French-learning infants completed the experiment. The stimuli, design and procedure were identical to those of Experiment 1.

3.2. Results

As in Experiment 1, infants’ looking time to the grammatical trials ($M=42.69$ sec; $SE=5.06$ sec) was not different from that of the ungrammatical trials ($M=42.51$ sec; $SE=4.73$ sec), $t(13)=0.052$, $p=.959$, 2-tailed. Like the 14-month-olds, the 17-month-old infants showed no evidence of categorization when only the determiner cue was present.

![Fig. 2. 17-month-old infants’ looking (listening) times (means and standard errors) to the two test trial types, grammatical versus ungrammatical.](image)

4. Experiment 3

With the current design, both 14- and 17-month-olds failed to show evidence of categorization based on purely distributional grounds. These results differ from Höhle, et al. (2004), Mintz (2006), Gomez et Lakusta (2004) and Gerken, et al. (2005), who showed categorization at this age or even slightly younger. It is possible that infants of this age indeed cannot do purely distributionally based categorization. Nevertheless, the null results could also be due to the new method used in the present study. Previous published studies using this preferential procedure typically used still images as stimuli. Our study was the first to use an animated talking puppet as the visual presentation. To rule out the possibility that this method may not be reliable, we conducted the same experiment with much older children, who should definitely have the ability to categorize grammatical gender. We predicted that older children would show a looking difference for grammatical versus ungrammatical trials.

4.1. Participants, Stimuli, Design and Predictions
Participants were twelve 30-month-old (Mean age = 935 days; SD = 11.19; range: 917-959) monolingual Quebec-French-learning infants. The stimuli, design and procedure were identical to those of Experiments 1 and 2.

4.2. Results

The same analysis as Experiment 1 and 2 were conducted on infants’ looking data. As predicted, there was a significant difference between the looking times for the two trial types, $t(11) = -3.024$, $p = .12$, 2-tailed. Infants looked significantly longer while listening to ungrammatical trials (Mean = 39.14 sec; SE = 3.13 sec) as opposed to the grammatical trials (Mean = 48.18 sec; SE = 3.42 sec). That is, infants aged 30 months were able to perform gender categorization. The results also suggest that the new method was reliable.

![Fig. 3. 30-month-old infants’ looking (listening) times (means and standard errors) to the two test trial types, grammatical versus ungrammatical.](image)

5. General Discussion

The interest of the present study was to determine if infants can categorize novel words based on contextual distributional cue alone. In previous studies infants aged between 12 and 17 months showed evidence of categorizing nouns (Höhle, et al., 2004), verbs (Mintz, 2006), grammatical genders (Gerken, et al., 2005), and content-word-like categories in an artificial language (Gomez & Lakusta, 2004). Those studies all used non-words as the target words to be categorized, and the stimuli contained no referential context. Therefore, no semantic information was present that could have assisted infants’ performance. Categorization was driven either by contextual distributional cues alone, or by phonological cues alone, or by the association of these two types of information.

In our present study we used a well-controlled design to test infants’ distributionally based categorization. Grammatical genders in French were used as the testing case. Pseudo-nouns that contained no phonological cues to gender were determined through pre-experimentation. In addition, any potential
acoustic and prosodic cues to gender were removed by using cross-spliced gender neutral productions of the pseudo-nouns. Thus, the only remaining information was the contextual distributional patterns between the pseudo-nouns and the adjacent gender specific determiners.

We found that both 14- and 17-month-old infants failed to perform gender categorization in our task. On the other hand, 30-month-olds showed robust evidence of categorization. Therefore, our results suggest that the ability to perform grammatical categorization based on distributional patterns alone, i.e., without requiring the nouns to contain any acoustic/prosodic or phonological cues, seems to be developed at some point between 17 and 30 months.

The lack of categorization in our 14- and 17-month-old infants contrasts with the results of Höhle, et al., (2004) and Mintz (2006) in which infants of this age showed evidence of categorization. The difference may be due to the fact that the stimuli in our study contained only the contextual distributional cues associated with a preceding function word, whereas the other studies contained fuller distributional information and possibly acoustic/prosodic cues. Our results are compatible with those of Gerken, et al. (2005) which also found that distributional cues alone did not lead to gender categorization.

Interestingly, in a recent study, 14–month-old French-learning infants categorized determiners as a class (Shi & Melançon, in press) when stimuli were also controlled against prosodic and phonological cues. It is possible that gender categorization occurs later in language development than a more global determiner categorization. However, caution should be taken in making such a conclusion. The task in the present study was more complex (involving more novel nouns and harder determiner pairings to learn) than that of the Shi and Melançon (in press) study. It is an empirical question whether gender categorization will be evident in 14-17-month-old infants in a simpler task as in Shi and Melançon (in press).

In addition, a crucial difference between the present study and that of Shi and Melançon (in press) lies in the design of stimuli. In Shi and Melançon (in press), infants were familiarized with pseudo-words preceded by determiners (or pronouns for another group of infants). After familiarization, infants were tested with the same pseudo-words preceded by different determiners versus by pronouns. In that design, infants had the option of not attending to the pseudo-words, although they may well do it. They could simply focus on the categorization of the function words across familiarization and test phases (i.e., which set of function words during the test can be grouped with the function words during the familiarization phase), without necessarily paying attention to any contextual distributional patterns with the content words. In the present study, infants were familiarized with two pseudo-words preceded by masculine determiner and two other pseudo-words preceded by a feminine determiner. After familiarization, infants were tested with the same pseudo-words preceded by different determiners, again masculine for two of the pseudo-words and feminine for the other two. That is, an infant would hear both kinds of determiner pairings during the familiarization and test phases. There was no
possibility of obtaining any positive results if infants were only attending to the
determiners from the familiarization to the test phases. Thus, unlike the Shi and
Melançon (in press) study, the present study was designed to specifically test
contextual distribution. In doing so, we did not dissociate the classification of
these two categories. Thus, positive results in the present study would imply
both noun gender categorization and determiner gender categorization. Our 30-
month-old infants indeed demonstrated this sophisticated ability. On the other
hand, null results in the present design would imply that infants failed to track
the distributional patterns of both categories.

Related to our study, it is interesting to consider the model proposed by
Braine (1987), which suggests that there are two levels of learning. In Level 1,
the association of function morpheme distribution with content word
phonological features is needed for categorization. Level 1 learning may involve
establishing the association of such phonological gender marking on the nouns
with the gender of the determiners. At Level 2, such association would be no
longer needed, and categorization can be based solely on function word
distribution with nouns. Level 1 learning is consistent with the results of Gomez
and Lakusta (2004) and Gerken et al. (2005), where infants were relying on the
association of specific functional morphemes and certain phonological cues on
the noun to form categories. The results of the 30-month-olds in the present
study may represent Level 2 knowledge.

From the syntactic perspective, categories are defined in terms of their
structural relations in syntax, independent of any phonological and acoustical
properties of the words under each category. In this sense, the ability to use
function words alone for categorization may reflect a higher level knowledge
than the categorization that relies on the association of function words with
certain characteristic phonological cues on the content words.

Our results suggest that infants may require the association of function word
distributional cues with content word phonological/acoustic/prosodic cues for
initial categorization, before being able to ultimately categorize without relying
on the association. In the case of the French gender system, phonologically
typical masculine and feminine cues on some nouns might enable infants to start
their initial categorization (e.g., the typical masculine nouns ending –eau can be
linked to masculine determiners). The outcome of this learning may enable
infants to perform distribution-only categorization at a later stage, which may
eventually guide them to reach the level of abstract, structurally based
categorization as defined for the mature grammar. In further research, we plan to
examine whether infants indeed need to associate some acoustic/prosodic or
phonological cues of the gender of nouns with the distribution of the determiners
as a first step for grammatical gender categorization.

Acknowledgment

This work was funded by grants from NSERC, SSHRC and CFI.
References


