



Infants' Acquisition of Grammatical Gender Dependencies

Marieke van Heugten and Anne Christophe
Laboratoire de Sciences Cognitives et Psycholinguistique
(ENS/EHESS/CNRS)
Département d'Études Cognitives
École Normale Supérieure – PSL Research University

To successfully understand spoken language, listeners need to determine how words within sentences relate to one another. Although the ability to compute relationships between word categories is known to develop early in life, little research has been conducted on infants' early sensitivity to subcategorical dependencies, such as those evoked by grammatical gender (where the article form is dictated by the noun's gender). This study therefore examines whether French-learning 18-month-olds track such relationships. Using the Visual Fixation Procedure, infants were presented with article–noun sequences in which the gender-marked article either matched (e.g., *la*_{FEM} *poussette*_{FEM} “the stroller”) or mismatched (e.g., *le*_{MASC} *poussette*_{FEM}) the gender of the noun. A clear preference for correct over incorrect co-occurrences was observed, suggesting that by 18 months of age, children's storage and access of words is sufficiently sophisticated to include the means to track subcategorical dependencies. This early sensitivity to gender information may be greatly beneficial for constraining lexical access during online language processing.

Within their first few years of life, children rapidly and without much instruction acquire their native language. By the time they reach their first birthday, infants typically start pronouncing their first words and their productive vocabulary quickly expands to an approximate average of 300

Correspondence should be sent to Marieke van Heugten, Department of Psychology, University at Buffalo, State University of New York, Buffalo, NY 14260. E-mail: mariekev@buffalo.edu

words around age two (Dale & Fenson, 1996). Over time, toddlers start to combine words into short phrases (e.g., Brown, 1973) and by the pre-school period, most children speak in full sentences. The building blocks for such language skills are, however, in place long before evidence for its productive use is observed. As early as 6 months of age, for instance, children map frequently occurring words in their language input onto visual representations of these items (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 2012) and in the ensuing months, infants start tuning into the formal relationships among linguistic elements as a function of the arrangement of words and phrases within an utterance.

Relationships between words within an utterance are often conveyed by means of function morphemes, such as determiners (e.g., *the*) or pronouns (e.g., *she*). Although these frequently occurring, closed-class words tend to be omitted from children's early productions and have traditionally been thought to be overlooked until long after children start talking (Brown, 1973), more recent work testing the perception of these items has revealed that children do process them. Toward the end of the first year of life, for example, infants encode and recognize the sound forms of function words (e.g., Shafer, Shucard, Shucard, & Gerken, 1998; Shi, Cutler, Werker, & Cruickshank, 2006) and use this knowledge to segment phrases into its individual components (Hallé, Durand, & de Boysson-Bardies, 2008; Shi & Lepage, 2008; Shi et al., 2006). Given that functors tend to be indicative of the word class of subsequent linguistic material (e.g., *the* is more likely to be followed by nouns such as *ball* than by verbs such as *walked*; Mintz, 2003; Redington, Chater, & Finch, 1998), knowledge of function words can furthermore constrain word candidates to the items of a specific syntactic category. Indeed, by 18 months of age, infants readily exploit function words both to deduce the word class of novel words (Höhle, Weissenborn, Kiefer, Schulz, & Schmitz, 2004; Shi & Melançon, 2010) and to facilitate the recognition of familiar words (Cauvet et al., 2014; Gerken & McIntosh, 1993; Kedar, Casasola, & Lust, 2006; Van Heugten & Johnson, 2011; Zangl & Fernald, 2007).

Relationships between words do not merely exist between major word classes, but can also involve more subtle subcategories. In French, for instance, the form of singular articles depends on the gender of the associated noun (e.g., *le*_{MASC} *doudou*_{MASC} “the security blanket” but *la*_{FEM} *poussette*_{FEM} “the stroller”). Although the dependency between article form and noun gender is largely independent of the semantics of the word,¹ French learners nonetheless exploit this co-occurrence pattern

¹French does have some semantically transparent words, mostly nouns referring to human entities (e.g., *la*_{FEM} *filles*_{FEM} ‘the girl’ vs. *le*_{MASC} *garçon*_{MASC} ‘the boy’), but these are considered the exception (Corbett, 1991).

during linguistic processing by 25 months of age (also see Johnson, 2005; Lew-Williams & Fernald, 2007; Van Heugten & Shi, 2009 for similar results with slightly older Dutch- and Spanish-learners). However, not much is known regarding the developmental trajectory of this sensitivity to grammatical gender. In particular, it is currently unclear whether children start acquiring this article–noun dependency early on in life or whether it is only learned around their second birthday, although there is reason to suspect that tracking gender co-occurrences can be challenging prior to age two. Specifically, artificial language work suggests that English-learning 17-month-olds gain sensitivity to an unfamiliar (though relatively simple) gender system only when multiple correlated phonological cues are available (Gerken, Wilson, & Lewis, 2005). In addition, acquiring the abstract grammatical categories of new words has been shown to be hard for young children when task demands are relatively high (Cyr & Shi, 2013). In that study, children were taught four novel word forms preceded by a gender-marked *indefinite* article as the sole cue denoting its gender (e.g., *un*_{MASC} *ravole*_{MASC}; “a <pseudoword>”). At test, children were presented with the same pseudowords preceded either by the correct or by the incorrect *definite* article (e.g., correct: *le*_{MASC} *ravole*_{MASC} “the <pseudoword>”; incorrect: *la*_{FEM} *ravole*_{MASC} “the <pseudoword>”). Using this design, children were found to differentiate between the two types of test trials only by 30 months of age, although 20-month-olds were subsequently shown to succeed when task demands were reduced.

What does this mean for learning gender dependencies in the case of naturalistic language input? On the one hand, children’s language input lacks the simplicity associated with laboratory experiments and learning gender from everyday input may be relatively hard compared to laboratory tasks. On the other hand, experiments typically last only a few minutes (compared to a more extensive time span available for learning dependencies in natural languages), potentially making it easier to learn novel dependencies from the input. To differentiate between these two possibilities, this study examines whether French-learning children track and store the subcategorical gender information of familiar words before their second birthday. Using the Visual Fixation Procedure, 18-month-olds were presented with lists of grammatical and ungrammatical noun phrases (NPs), the grammatical lists containing correct-gender definite articles (e.g., *le*_{MASC} *doudou*_{MASC} “the security blanket”; *la*_{FEM} *poussette*_{FEM} “the stroller”) and the ungrammatical lists containing incorrect-gender definite articles (e.g., *la*_{FEM} *doudou*_{MASC} “the security blanket”; *le*_{MASC} *poussette*_{FEM} “the stroller”). Following previous work pitting grammatical and ungrammatical utterances using similar procedures (e.g., Santelmann & Jusczyk, 1998; Van Heugten & Johnson, 2010; Van Heugten & Shi, 2010),

a preference for the correct over the incorrect NPs will be taken to indicate children's sensitivity to grammatical gender cues.

METHOD

Participants

A total of 24 typically developing monolingual French-learning 18-month-olds from the Paris area participated in this study (mean age: 18 months, 7 days; age range: 17 months, 21 days -19 months, 9 days; 12 girls). None of these children were known to have any language or hearing problem. An additional three infants were tested, but excluded from the analyses due to extreme fussiness. Participating children received a diploma as a token of appreciation.

Materials

Twelve nouns (six monosyllabic and six bisyllabic), generally known by 18-month-olds, were used in this study. Half of these nouns were masculine and the other half feminine (masculine: *chat* "cat", *chien* "dog", *pain* "bread", *bébé* "baby", *biberon* "bottle", *doudou* "security blanket"; feminine: *bouche* "mouth", *couche* "diaper", *main* "hand", *cuillère* "spoon", *compote* "stewed fruit", *poussette* "stroller"). A female native speaker of French recorded the materials. To avoid recording ungrammatical utterances (potentially resulting in unnatural tokens), the materials were cross-spliced. For each NP token used in the study, three versions were recorded. One version contained the noun preceded by the correct definite article (e.g., *la poussette* "the stroller"). From this version, the noun was spliced. In the other two versions, the noun was replaced by a word starting with the same consonant-vowel sequence. One of these nouns was masculine and the other one feminine. Both were preceded by the correct definite article (e.g., *le poulet* "the chicken"; *la poupée* "the doll"). From these two tokens, the article was selected. To prevent mismatches in co-articulation between the article vowel and noun onset, articles and nouns were spliced immediately following the noun onset consonant, just before the first vowel of the noun. The combination of the articles with the noun resulted in two NPs, one containing the correct definite article (e.g., *la poussette*) and one containing the incorrect definite article (e.g., *le poussette*), that were similar in duration and intonation (average correct NP length: 855 ms; average incorrect NP length: 865 ms).

Eight lists were created, four of which contained the correct NPs and four of which contained incorrect NPs. Lists contained two unique tokens

of each of the twelve NPs. These NPs were ordered differently across each of the four correct and incorrect lists, but the order was identical across these two conditions. Interstimulus pauses were approximately 750 msec long. All lists lasted 38.3 sec.

In addition to these auditory materials, which formed the core of the experiment, two movies were used. One movie, a silent clip from the Flurry screensaver featuring colorful swirls, was used to capture the children's looks toward the screen during stimulus presentation. The other movie, a red flashing light accompanied by a cartoon boing, was used as an attention getter during the intertrial period.

Procedure

Infants were individually tested for their spontaneous listening preferences using the Visual Fixation Procedure. During test, they sat on the parent's lap in front of a TV screen in a sound-attenuated booth. At the front of the booth, next to the TV, two loudspeakers were positioned at the child's ear level. Stimuli were presented using the Lincoln Infant Lab Package 1.0 software package (Meints & Woodford, 2008). The procedure started by displaying the attention getter on the screen. Once the child oriented toward the screen, the experimenter, monitoring the child's looking behavior through a video camera, initiated the first trial. During each trial, the experimenter pushed a button when the child oriented toward the screen. When the child looked away, the experimenter released the button. If the child reoriented toward the screen within 2 sec, the trial continued to play, but the time looked away was subtracted from the orientation time. Trials lasted until the child looked away for more than 2 sec or until the maximum list length was reached. The experiment finished when the child had been presented with all eight lists. All children listened to the same lists, although the order in which these lists appeared was randomized separately for each child.

RESULTS

Parental reports confirmed that children generally understood most of the words used in the study (mean number of words comprehended: 10.88 of 12; range: 6–12). Children oriented toward the screen for an average of 21.76 sec in correct trials and for an average of 17.59 sec in incorrect trials (see Figure 1), with 18 out of 24 children listening longer to correct than to incorrect trials. A two-tailed paired samples *t*-test revealed that these two values differ significantly from each other (4.17 sec; *SEM* = 1.52 sec; *t*

(23) = 2.747; $p = 0.011$; $d = 0.581$). Neither the child's age nor the number of test words parents reported their child to know correlated with this difference score. Thus, by 18 months of age, infants display a robust preference for correctly used over incorrectly used definite articles, indicating that they track the co-occurrence between article and noun during language processing.

DISCUSSION

This study shows that French-learning 18-month-olds track (and store) the subtle co-occurrence patterns between article form and noun gender. By presenting infants with correct and incorrect article–noun sequences, we identified a listening preference for correct over incorrect phrases, suggesting that children are sensitive to gender dependencies in the months prior to their second birthday. Children's daily language input thus provides them with the information necessary to learn subcategorical relationships early in life.

How could this early sensitivity to gender cues help children process spoken language? Although the current study does not explicitly test whether the NPs were segmented into its individual components, infants' early sensitivity to the form and usage of function words (Hallé et al., 2008; Shi & Lepage, 2008; Shi et al., 2006) makes it unlikely that 18-month-olds consider article–noun sequences to be unanalyzed chunks

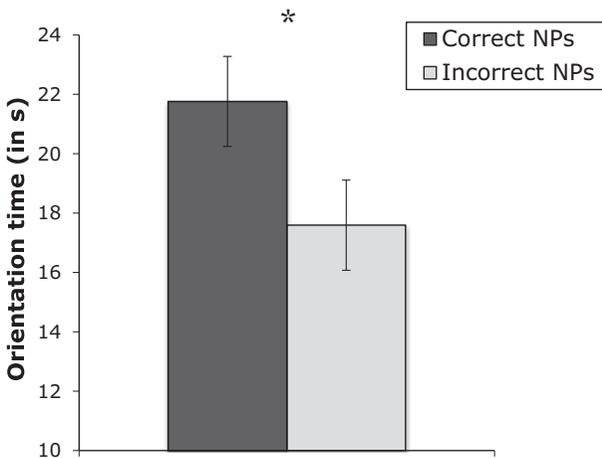


Figure 1 Mean orientation times in seconds (error bars indicate the standard error of the mean difference score) to correct and incorrect NPs.

(e.g., *lapoussette*). Instead, the recognition of an article (e.g., *la*) will allow children to infer the presence of a noun (Cauvet et al., 2014; Höhle et al., 2004; Shi & Melançon, 2010), and its associated gender cues may subsequently prevent (or eliminate) the activation of candidates that do not belong to the same subcategorical gender category. For example, upon hearing *la pou*, children may continue to consider both feminine-gender *poussette* “stroller” and *poupée* “doll” as potential nouns, but masculine-gender *poulet* “chicken” should no longer be an option (cf. Johnson, 2005; Lew-Williams & Fernald, 2007; Van Heugten & Shi, 2009 for similar patterns with older children). Note that the current study does not test whether the acquired gender knowledge is grammatical (i.e., masculine nouns should be preceded by masculine articles; feminine nouns should be preceded by feminine articles) or distributional (i.e., *doudou* “security blanket” should be preceded by *le*; *poussette* “stroller” should be preceded by *la*; though see Cyr & Shi, 2013 for evidence for grammatical abstraction). It may thus be either the surface form of the articles or their associated gender cues that could drive the potential advantage for online language processing. Future work could tease apart these two possibilities.

In sum, this study reveals that infants as young as 18 months track the co-occurrence patterns between articles and nouns in their daily language input in a robust fashion and use this information during speech perception. This early sensitivity to such a subtle subcategorical cue could be greatly beneficial for constraining lexical access and underlines the tremendous sophistication of early syntactic processing.

ACKNOWLEDGMENTS

This work was supported by a Fyssen Postdoctoral Grant, as well as grants from the French Agence Nationale de la Recherche (ANR-2010-BLAN-1901, ANR-13-APPR-0012, ANR-10-0001-02 PSL* and ANR-10-LABX-0087), the Fondation de France, and the Région Ile-de-France. We thank all children and their families for their participation.

REFERENCES

- Bergelson, E., & Swingle, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America*, 109(9), 3253–3258.
- Brown, R. (1973). Development of the first language in the human species. *American Psychologist*, 28(2), 97–106.

- Cauvet, E., Limissuri, R., Millotte, S., Skoruppa, K., Cabrol, D., & Christophe, A. (2014). Function words constrain on-line recognition of verbs and nouns in French 18-month-olds. *Language Learning and Development*, 10(1), 1–18.
- Corbett, G. G. (1991). *Gender*. Cambridge, UK: Cambridge University Press.
- Cyr, M., & Shi, R. (2013). Development of abstract grammatical categorization in infants. *Child Development*, 84(2), 617–629.
- Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, 28(1), 125–127.
- Gerken, L., & McIntosh, B. J. (1993). Interplay of function morphemes and prosody in early language. *Developmental Psychology*, 29(3), 448–457.
- Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language*, 32(2), 249–268.
- Hallé, P. A., Durand, C., & de Boysson-Bardies, B. (2008). Do 11-month-old French infants process articles? *Language and Speech*, 51(1-2), 23–44.
- Höhle, B., Weissenborn, J., Kiefer, D., Schulz, A., & Schmitz, M. (2004). Functional elements in infants' speech processing: The role of determiners in the syntactic categorization of lexical elements. *Infancy*, 5(3), 341–353.
- Johnson, E. K. (2005). Grammatical gender and early word recognition in Dutch. In A. Brugos, M. R. Clark-Cotton & S. Ha (Eds.), *Proceedings of the 29th annual Boston University Conference on Language Development* (Vol. 1, pp. 320–330). Somerville, MA: Cascadia Press.
- Kedar, Y., Casasola, M., & Lust, B. (2006). Getting there faster: 18- and 24-month-old infants' use of function words to determine reference. *Child Development*, 77(2), 325–338.
- Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, 18(3), 193–198.
- Meints, K., & Woodford, A. (2008). Lincoln Infant Lab Package 1.0: A new programme package for IPL, Preferential Listening, Habituation and Eyetracking. [WWW document: Computer software & manual]. Retrieved from <http://www.lincoln.ac.uk/psychology/baby-lab.htm>
- Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90(1), 91–117.
- Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, 22(4), 425–469.
- Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition*, 69(2), 105–134.
- Shafer, V. L., Shucard, D. W., Shucard, J. L., & Gerken, L. (1998). An electrophysiological study of infants' sensitivity to the sound patterns of English speech. *Journal of Speech, Language and Hearing Research*, 41(4), 874–886.
- Shi, R., Cutler, A., Werker, J., & Cruickshank, M. (2006). Frequency and form as determinants of functor sensitivity in English-acquiring infants. *The Journal of the Acoustical Society of America*, 119(6), EL61–EL67.
- Shi, R., & Lepage, M. (2008). The effect of functional morphemes on word segmentation in preverbal infants. *Developmental Science*, 11(3), 407–413.
- Shi, R., & Melançon, A. (2010). Syntactic categorization in French-learning infants. *Infancy*, 15(5), 517–533.
- Tincoff, R., & Jusczyk, P. W. (2012). Six-month-olds comprehend words that refer to parts of the body. *Infancy*, 17(4), 432–444.
- Van Heugten, M., & Johnson, E. K. (2010). Linking infants' distributional learning abilities to natural language acquisition. *Journal of Memory and Language*, 63(2), 197–209.

- Van Heugten, M., & Johnson, E. K. (2011). Gender-marked determiners help Dutch learners' word recognition when gender information itself does not. *Journal of Child Language*, 38(1), 87–100.
- Van Heugten, M., & Shi, R. (2009). French-learning toddlers use gender information on determiners during word recognition. *Developmental Science*, 12(3), 419–425.
- Van Heugten, M., & Shi, R. (2010). Infants' sensitivity to non-adjacent dependencies across phonological phrase boundaries. *The Journal of the Acoustical Society of America*, 128(5), EL223–EL228.
- Zangl, R., & Fernald, A. (2007). Increasing flexibility in children's online processing of grammatical and nonce determiners in fluent speech. *Language Learning and Development*, 3(3), 199–231.