J. Child Lang. **38** (2011), 87–100. © Cambridge University Press 2010
doi:10.1017/S0305000909990146

Gender-marked determiners help Dutch learners' word recognition when gender information itself does not*

MARIEKE VAN HEUGTEN AND ELIZABETH K. JOHNSON
University of Toronto, Mississauga, Canada

(*Received* 16 *February* 2008 – *Revised* 19 *September* 2008 – *Accepted* 19 *June* 2009 – *First published online* 25 *January* 2010)

ABSTRACT

Dutch, unlike English, contains two gender-marked forms of the definite article. Does the presence of multiple definite article forms lead Dutch learners to be delayed relative to English learners in the acquisition of their determiner system? Using the Preferential Looking Procedure, we found that Dutch-learning children aged 1;7 to 2;0 use articles during sentence comprehension in a fashion comparable to similarly aged English learners. That is, Dutch learners' sentence processing was impaired when a nonsense (*se*) as opposed to real article (*de*, *het*) preceded target words, much like English learners' sentence processing is disrupted by the use of a nonsense article. At the same time, however, gender cues did not help Dutch learners recognize target nouns more efficiently, indicating that gender has yet to be acquired. Thus, although Dutch-learning children aged 1;7 to 2;0 have not mastered all aspects of their language's article system, they nonetheless use their partial knowledge of articles during speech processing.

Early theories of language acquisition were based primarily on corpus studies, and were therefore focused on explaining children's production data. Since function words (e.g. articles such as *the* and *a*) are typically

[*] We would like to thank all the families that participated, Angela Khadar, Margret van Beuningen, Sara Bögels, Ellen Westrek and Monique Diks for their assistance with recruitment, preparing, running and coding this study, as well as Joost van de Weijer for the use of his corpus. This research was funded by an NWO Spinoza Grant entitled 'Native and Non-native Listening', awarded to Anne Cutler. Parts of this research were presented at the 12th Annual Conference on Architectures and Mechanisms for Language Processing (Nijmegen, the Netherlands, August 2006) and at the 31st Boston University Conference on Language Development (Boston, MA, November 2006). Address for correspondence: Marieke van Heugten, Department of Psychology, University of Toronto Mississauga, 3359 Mississauga Road N., Mississauga, ON, L5L 1C6, Canada. e-mail: marieke.vanheugten@utoronto.ca

omitted from children's early utterances, some language researchers reasoned that these short, acoustically reduced items are not among the earliest words acquired by children (e.g. Brown, 1973). In fact, it was argued that children begin speaking before having developed any knowledge of free-standing function words. This apparent late acquisition of function words was hypothesized to be due to their lack of saliency in fluent speech (e.g. Echols & Newport, 1992; Gleitman & Wanner, 1982). Since function words are crucial for conveying the grammatical structure of an utterance, the notion that young toddlers do not attend to these words strongly impacted early theories of language acquisition.

Experimental studies, however, have shed a different light on function word acquisition. Not only do we now know that children start learning function words before they speak, the acquisition of these terms is thought to play an important role in enabling infants to bootstrap the phonological structure of their language (e.g. Christophe, Millotte, Bernal & Lidz, 2008; Morgan, Shi & Allopenna, 1996). That is, function words appear to be one of the first, rather than one of the last things children become sensitive to in their language.

Three lines of evidence were particularly influential in driving this revised view of function word acquisition. First, infants are attentive to the phonetic detail of unstressed syllables in fluent speech (e.g. Johnson, 2005a), which undermines the view that toddlers do not produce function words because they have difficulty perceiving them. Second, speech imitation tasks have demonstrated that toddlers' knowledge of function words influences their omission patterns. English learners, for example, are more likely to omit existing function words as opposed to acoustically reduced nonsense items, indicating that they possess fine-grained knowledge about the typical form of function words (Gerken, Landau & Remez, 1990). Third, and perhaps most convincingly, comprehension studies have demonstrated that two-year-olds, who do not yet reliably produce function words, nonetheless follow instructions better when real (e.g. *Find the ball for me*) as opposed to nonsense or ungrammatically positioned function words (e.g. *Find gub/was ball for me*) are used (Gerken & McIntosh, 1993; Shipley, Smith & Gleitman, 1969), thus showing children's use of article information during sentence comprehension. More recently, these results have been replicated with toddlers as young as 1;6 using the Preferential Looking Paradigm (Kedar, Casasola & Lust, 2006; Zangl & Fernald, 2007).

Here, we will examine function word acquisition in Dutch. While the Dutch and English article systems are highly similar in many aspects (e.g. obligatoriness and consistency of definite articles), they differ in that Dutch marks definite articles for gender. In Dutch, common gender nouns are preceded by the definite article *de* (e.g. *de bal* 'the$_{\text{COMMON}}$ ball'), while neuter gender nouns are preceded by the definite article *het* (e.g. *het boek*

'the$_{\text{NEUTER}}$ book'). The presence of two forms of the definite article may complicate article acquisition for at least two reasons. First, Dutch children need to learn that although there is only one form of the indefinite article, there are two allomorphs of the definite article. That is, they need to observe and accept that two forms have the same meaning and occur in the same syntactic position (e.g. Clark, 1993). Thus, even though some researchers may argue that this increased paradigmatic complexity of the Dutch determiner system would facilitate acquisition, we adopt the view that this greater complexity may in fact slow determiner acquisition in Dutch relative to English. Second, it stands to reason that in any language containing multiple forms of the definite article, the frequency of each form is reduced compared to the frequency of the definite article in languages containing only one definite article. In order to confirm this for Dutch versus English, we used the Childes Corpus (MacWhinney, 2000) to compare the parental input for seven Dutch-learning (the Groningen corpus; Wijnen & Verrips, 1998; 305,413 words in our sample) and six American English-learning children (the Providence corpus; Demuth, Culbertson & Alter, 2006; 958,413 words) between the ages of 1;06.00 and 3;00.00. The relative frequency of each of the Dutch definite articles (0·022 for *de* and 0·022 for *het*[1]) as a function of the total number of words was found to be lower than the relative frequency of English *the* (0·034). This reduced frequency of the individual definite article forms may also delay determiner acquisition, since less frequent function words are known to be acquired later than more frequent function words (e.g. Shi, Cutler, Werker & Cruickshank, 2006; Shi & Lepage, 2008). Thus, for both reasons, article acquisition in Dutch may lag behind that in English.

Indeed, corpus studies suggest that Dutch learners may be somewhat slower than English learners in the acquisition of determiners (e.g. Rozendaal & Baker, 2008). Production data, however, do not allow us to determine if a delay is already present before children start speaking. An interesting question thus concerns what children know about function words before they start producing them. In this study, we begin to examine this by testing Dutch-learning toddlers' use of articles during sentence comprehension. We chose to test children aged 1;7 and 2;0 because previous Preferential Looking studies have suggested that French and English learners start using definite articles during word comprehension sometime between 1;6 and 2;1 (Kedar *et al.*, 2006; Van Heugten & Shi, 2009; Zangl & Fernald, 2007), leading to the prediction that in the latter half of their

[1] Note that *het* can sometimes take on other grammatical functions such as expletive pronoun (e.g. *het regent* 'it rains') or provisional subject (e.g. *het is goed dat je dat gedaan hebt* 'it is good that you did that'). Since we did not differentiate between these different functions, our count may have artificially inflated the frequency of the definite article *het*.

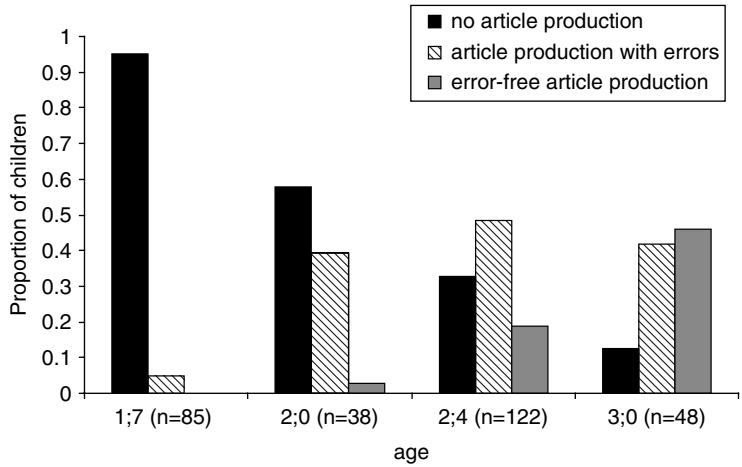

Fig. 1. In a parental report survey, we asked parents to indicate whether their children have started producing definite articles (i.e. *de* and *het*), and if they have, we asked how adult-like these productions are. Parents were given three options to choose from: no article production, article production with errors or error-free article production. Article errors could either be overgeneralization mistakes where a neuter gender word is preceded by a common gender article (e.g. *de boek* 'the$_{\text{COMMON}}$ boek$_{\text{NEUTER}}$') or the production of filler determiners where the child's realization of the article is not the same as the target determiner. This provides us with a rough approximation of children's productions of definite articles. The bars represent the proportion of children at different ages that (according to parental report) are still at the bare noun stage (no article production), that have started producing definite articles but still make frequent mistakes (article production with errors), and that reliably and correctly produce definite articles (error-free article production).

second year of life, Dutch learners would also begin using function words to comprehend sentences. Further support for this notion comes from a lab survey we carried out with Dutch parents. According to parental report, Dutch learners below the age of 1;8 typically do not produce any definite articles. By 2;0, however, toddlers tend to start producing *de* and *het* (albeit with frequent errors; see Figure 1; see also Rozendaal & Baker, 2008). This shift in the productive use of articles may reflect a developmental change in the way children process articles. Were this to be the case, we would expect to see differences between children aged 1;7 and 2;0 in their article–noun integration.

Children were tested in the Preferential Looking Procedure, modeled after previous perceptual studies examining article acquisition (Johnson, 2005b; Kedar *et al.*, 2006; Lew-Williams & Fernald, 2007; Van Heugten & Shi, 2009; Zangl & Fernald, 2007). On each trial, children heard sentences asking them to look at one of two depicted objects on a TV screen. Target words were either preceded by the correct or a nonsense definite article. Nonsense articles are the strongest test case of receptive

article use as they resemble proto-determiners, or schwa-like realizations of determiners, produced during the early stages of language production (e.g. Taelman, Durieux & Gillis, 2009; Rozendaal & Baker, 2008). The presence of these forms in Dutch learners' early speech may indicate that they know determiner slots should be filled with a reduced syllable. Alternatively, Dutch learners may process more fine-grained phonological detail before this becomes apparent in their own productions. If Dutch learners begin using definite articles to recognize upcoming nouns at the same age as English learners, Dutch learners aged 1;7 and 2;0 should recognize target nouns more efficiently during correct trials than during nonsense trials, where the typical article–noun combination is violated. However, if Dutch learners' use of *de* and *het* is delayed relative to English learners' use of the definite article *the*, then only the older (or neither) group of Dutch learners may show this effect. Correct trials were further subdivided such that the words referring to the two pictures on the screen could either be of different or of the same gender (cf. Johnson, 2005b; Lew-Williams & Fernald, 2007; Van Heugten & Shi, 2009). The article disambiguated between the two depicted nouns in the different- but not the same-gender correct trials. Evidence for sensitivity to the gender system would thus be obtained if target nouns are recognized more efficiently in different- compared to same-gender correct trials. A previous Preferential Looking study with Dutch learners, however, has suggested that this ability only begins to appear after age 2;0 (Johnson, 2005b; Johnson & Diks, 2005).

METHOD

Participants

Thirty-two normally developing Dutch-learning children aged 1;7 (age range: 1;06.05–1;07.28, 15 boys) and 32 children aged 2;0 (age range: 2;00.03–2;00.29, 18 boys) were tested in this study. An additional eight children aged 1;7 and eight aged 2;0 were tested, but excluded from the sample prior to data coding due to fussiness or failure to finish the study.

Stimuli

A total of eight plosive-initial target nouns, generally known by children aged 1;6,[2] were chosen as target words. Half of these items were *de*-words (*koe* 'cow', *kip* 'chicken', *bal* 'ball', *buik* 'belly') and half were *het*-words (*kind* 'child', *konijn*, 'rabbit', *boek* 'book' and *been*, 'leg'). In addition to the

[2] Nouns were considered appropriate if parents of Dutch learners aged 1;5 who had previously visited our lab and who had filled out the Dutch version of the CDI indicated that their child understood these words. All target nouns were judged to be familiar for at least 70% of the children aged 1;5.

eight target nouns, four filler nouns (two *de*- and two *het*-words) were selected. For each target and filler noun, a visual image was selected to represent the word. Images were brightly colored and approximately equal in size.

Target and filler nouns were embedded in sentences (*Zie je [article] [noun]?* 'Do you see [article] [noun]?'). The nonsense syllable *se* was chosen to serve as the article in the nonsense sentences as it shares vowel properties with both the article *de* and the reduced form of the article *het* ([ət], the form that is most commonly used in casual speech). Thus, unlike studies with English learners that used full syllables (e.g. *loo shoe*; Gerken & McIntosh, 1993; Zangl & Fernald, 2007), the nonsense syllable in this study is acoustically similar to real function words.[3] All test sentences were followed by a positive comment about the pictures (e.g. *Wat leuk, zeg!* 'How nice!'). In order to increase variability in our materials and hence prevent infants from employing laboratory-induced strategies, we included filler trials with different sentence structures (e.g. *Kun je de schoen vinden* 'Can you find the shoe?'). Filler trials always used the appropriate article (i.e. *de* or *het*) and each article was presented equally often in the experiment.

A female native Dutch speaker, naive to the purpose of the study, recorded the utterances in a child-directed manner (sample rate: 44·1 kHz, 16 bits). In contrast to previous studies (Kedar *et al.*, 2006; Zangl & Fernald, 2007), the sentences used in both correct and nonsense trials were

[3] In certain regions of Holland (including Nijmegen), speakers typically devoice word-initial fricatives. For this reason, our speaker's *se* was acoustically very similar to the pronoun *ze* 'she'. Given that the children tested in our study were all born in Nijmegen, it is likely that a good portion of the children we tested did not distinguish *se* from *ze*. For these children, this experiment would test their detection of ungrammatically positioned function words (e.g. look at *she* ball) rather than detection of nonsense function words (look at *el* ball). Although we did not test whether children in this study interpreted *se* as *ze*, this does not change the question of whether reduced syllables in determiner position are fully encoded and used when comprehending speech.

As pointed out by an anonymous reviewer, the presence of a nonsense article acoustically similar to a pronoun may raise the question of whether our results reflect early determiner or pronoun knowledge. That is, children's recognition of target nouns could have been slowed in the nonsense article condition because a verb (rather than a noun) was expected to follow *se*. While we do not want to exclude the possibility that the children in our study had pronoun knowledge, it seems unlikely that children ONLY use pronoun knowledge (and no determiner knowledge) during word comprehension. In fact, a corpus count using the same Dutch Groningen corpus revealed that both *de* and *het* are more frequent in infant-directed speech than *ze*. Given that frequency of the individual function word plays a crucial role in its acquisition (Shi *et al.*, 2006; Shi & Lepage, 2008), with more frequently occurring items being acquired earlier than less frequently occurring items, it seems more plausible that knowledge of *de* and *het* precedes knowledge of *ze*. This is corroborated, at least for German (Höhle *et al.*, 2004). In this study, infants were found to anticipate the presence of a noun following a determiner but not the presence of a verb after a pronoun.

created through cross-splicing to ensure that all articles were produced equally naturally. For all conditions, the same token of the target noun was used, while the preceding sentence was varied in order to create the different conditions. The token of the noun (e.g. *boek* 'book') was spliced from a grammatical instance of the sentence (e.g. *Zie je het boek?* 'Do you see the book?'). This noun token was then attached to a (preceding) carrier sentence. For sentences with nonsense articles, proper names starting with an unstressed *se* were used to precede the noun (e.g. *Zie je Sebastiaan?* 'Do you see Sebastiaan?'). For the sentences with correct articles, the carrier sentence was constructed from a sentence containing the correct article, the target noun and a bisyllabic verb (e.g. *Zie je het boek liggen?* 'Do you see the book lying?'). Note that the bisyllabic verb was added to control for the number of post-article syllables between the nonsense and correct sentence recording material (i.e. *Sebastiaan* vs. *het boek liggen*). Only the initial part of the sentence up until the article (e.g. *Zie je se* 'Do you see [nonsense article]'; *Zie je het* 'Do you see the') was used to create the actual stimuli and hence neither the proper name (in the nonsense article condition) nor the noun and verb (in the correct article conditions) appeared in the final materials. The carriers and targets were then combined to form grammatically correct (e.g. *Zie je het boek?* 'Do you see the book?'; *Zie je de bal?* 'Do you see the ball?') and grammatically incorrect (nonsense) sentences (e.g. *Zie je se boek/bal?* 'Do you see [nonsense article] book/ball?'). Sentences with correct articles lasted 1·055 s ($SD = 0{\cdot}114$) on average, while sentences with nonsense articles had a mean length of 1·093 s ($SD = 0{\cdot}107$). Target word duration was identical across conditions, as the same noun token was used.

Design

Three experimental trial types were created. During correct trials, toddlers heard instructions containing the correct definite article (e.g. *Zie je de bal?* 'Do you see the ball?'). The gender of the two nouns represented by the pictures on the TV screen could either be the same (e.g. a $\text{ball}_{\text{COMMON}}$ and a $\text{belly}_{\text{COMMON}}$; same-gender correct trials) or different (e.g. a $\text{ball}_{\text{COMMON}}$ and a $\text{book}_{\text{NEUTER}}$; different-gender correct trials). During nonsense trials, in which the two pictures always represented nouns of different gender, toddlers heard instructions containing the nonsense article *se* (e.g. *Zie je se bal?* 'Do you see [nonsense article] ball?'; see Table 1).

Toddlers were randomly assigned to one of four test lists. Each child was presented with trials of all conditions: four different-gender correct trials, four same-gender correct trials, four nonsense trials and eight fillers. Thus, only one-fifth of the trials contained a nonsense article. Half of the targets in each trial type were common gender and half were neuter gender. To avoid

TABLE 1. *Overview of the three experimental conditions*

Condition	Picture 1	Picture 2	Audio
Different-gender grammatical trials	*bal* ball$_{\text{COMMON}}$	*boek* book$_{\text{NEUTER}}$	*Zie je de bal* (*het boek*)*?* 'Do you see the ball (the book)?'
Same-gender grammatical trials	*bal* ball$_{\text{COMMON}}$	*buik* belly$_{\text{COMMON}}$	*Zie je de bal* (*de buik*)*?* 'Do you see the ball (the belly)?'
Nonsense trials	*bal* ball$_{\text{COMMON}}$	*boek* book$_{\text{NEUTER}}$	*Zie je se bal* (*se boek*)*?* 'Do you see [se] ball ([se] book)?'

confusion, within each of the four pseudo-randomized orders, a given picture was always presented on the same side of the screen. A ball, for example, would always appear on the left side of the screen for one child. For another child, it would always appear on the right. Within each order, targets occurred equally on left and right.

Procedure

Children were seated on their parent's lap approximately one meter in front of a 50 in TV screen. On each trial, two pictures were simultaneously presented side by side on the TV for a total of seven seconds. The pictures gradually increased in size to attract the toddlers' attention. The target noun started four seconds after picture onset. One picture always matched the target word in the instruction. On all trials, target and distracter had the same onset (e.g. *bal* 'ball' and *book* 'book'). A colorful ball accompanied by a short cartoon 'boing' was presented between trials. To prevent biases, the parent listened to masking music over headphones. The videos lasted approximately three minutes.

Off-line coding

Sessions were videotaped and subsequently coded off-line using Super-Coder. For each frame in a trial, a coder judged whether the toddler was looking left, right or at neither picture. The coder was always blind to the condition of a trial. Eight sessions (four for each age group) were randomly selected to be recoded by a second coder. The agreement between the two coders was consistently high (mean correlation = 0·98; $SD = 0{\cdot}046$).

RESULTS

Following precedents set by Preferential Looking studies with similarly aged children (e.g. Zangl & Fernald, 2007; Van Heugten & Shi, 2009) we used the proportion of looking time (LT) to target over a two-second period

as the dependent variable. This proportion of LT to target was calculated as the total LT to target divided by the total LT to target and the distracter. If a child, for example, looked to the target for 900 ms and to the distracter for 600 ms, the proportion of LT to target would be $900/(900+600)=0{\cdot}60$. Looks away from the screen were not taken into account. LTs were computed both before and after target onset. The LTs before onset were only used in the first analysis, where we compared the LTs to target before with the LTs to target after target onset. This was done to ensure that children recognized the words. LTs after target onset were also used to test for differences between the various conditions. We started measuring 320 ms after target onset as eye movements occurring before that time were likely initiated before target word onset (e.g. Lew-Williams & Fernald, 2007; Van Heugten & Shi, 2009; Zangl & Fernald, 2007).

First, we examined whether infants recognized the target words. The mean proportion of LT to target in the analysis windows was 0·50 (SEM = 0·009) before and 0·62 (SEM = 0·014) after target onset. A 2 (time; before and after) × 2 (age; 1;7 and 2;0) ANOVA showed that Time was indeed significant ($F(1, 62)=69{\cdot}656$; $p<0{\cdot}001$, $\eta_p^2=0{\cdot}53$). Thus, children looked longer to the target picture during the two-second time window following than during the two-second time window preceding target onset, showing that they recognized the targets. No other main effect or interaction was found.

We then tested the effect of Condition. A 3 (condition; same-gender correct, different-gender correct and nonsense) × 2 (gender; common and neuter) × 2 (age; 1;7 and 2;0) ANOVA was carried out on the proportion of LT to target during the two-second analysis window after target word onset. A main effect of Gender ($F(1, 62)=20{\cdot}446$; $p<0{\cdot}001$, $\eta_p^2=0{\cdot}25$) was found, indicating that common gender nouns were looked at more than neuter gender nouns. There was also a main effect of Condition ($F(1, 61)=6{\cdot}598$, $p=0{\cdot}003$, $\eta_p^2=0{\cdot}18$). Importantly, however, no other main effect or interactions reached significance. In other words, neither the gender of the target word nor toddlers' age modulated the Condition effect. Thus, the pattern of results does not differ between common and neuter gender words nor between children aged 1;7 and 2;0. For this reason, age and gender were collapsed in subsequent analyses. The overall effect of Condition held for both common and neuter gender nouns ($F(2, 62)=4{\cdot}031$; $p=0{\cdot}023$, $\eta_p^2=0{\cdot}12$; $F(2, 62)=3{\cdot}907$, $p=0{\cdot}025$, $\eta_p^2=0{\cdot}11$, respectively).

The average proportion of looking time to target was 0·65 (SEM = 0·018) for different-gender correct trials, 0·65 (SEM = 0·022) for same-gender correct trials and 0·56 (SEM = 0·020) for incorrect trials (see Figure 2). In line with the results found in Johnson & Diks (2005), planned comparisons showed no difference between different- and same-gender correct trials ($F(1, 63) < 1{\cdot}0$). This suggests that Dutch learners aged 1;7 to 2;0 have not

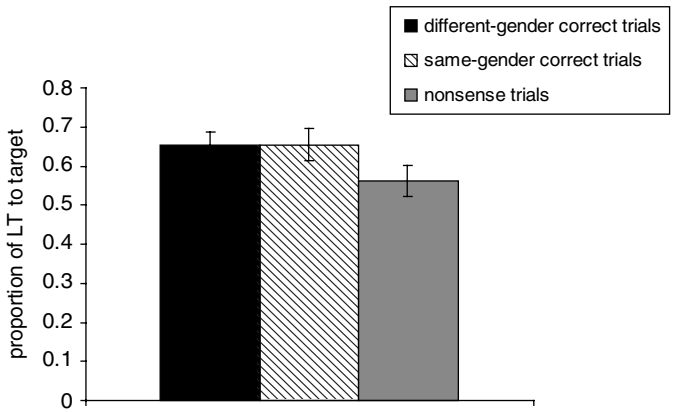

Fig. 2. Mean proportion of looking time to target over a two-second window for each condition (error bars indicate confidence intervals).

yet learned which article should precede which noun in the target language. In contrast, the difference between each of the correct instructions and the incorrect instruction was significant ($F(1, 63) = 14{\cdot}199$, $p < 0{\cdot}001$, $\eta_p^2 = 0{\cdot}18$; $F(1, 63) = 9{\cdot}971$, $p = 0{\cdot}002$, $\eta_p^2 = 0{\cdot}14$ for different- and same-gender comparisons, respectively), indicating that words were recognized more efficiently when real as opposed to nonsense function words preceded the targets. Thus, at an age when gender-marking information on determiners has yet to be acquired, Dutch learners nevertheless use determiners in speech comprehension. In other words, gender-marked determiners help Dutch learners' word recognition when gender information itself does not.

DISCUSSION

Our results demonstrate that Dutch-learning children aged 1;7 to 2;0 comprehend utterances better when they contain real as opposed to nonsense articles. Thus, despite the increased complexity of the article system in Dutch relative to English, Dutch learners are similar to English learners in that they begin using these forms in word comprehension well before their second birthday.[4] This finding is striking given children's failure to reliably produce definite articles at this age. Dutch learners' production patterns at this age are characterized by the presence of acoustically reduced proto-determiners, mostly realized as a schwa (e.g. Rozendaal & Baker, 2008; Taelman *et al.*, 2009). The results of our study suggest that this lack

[4] Whether this effect is grammatical in nature or whether it is a matter of local co-occurrences cannot be determined based on our results and is a question beyond the scope of this paper (see Van Heugten & Shi, 2009, for discussion).

of specificity in early production does not reflect the phonological detail with which definite articles are processed. In short, Dutch toddlers process definite articles with more phonological detail than their production pattern reveals.

An additional contribution of this study is the replication of earlier work suggesting that Dutch-learning children aged 2;0 have not yet acquired the gender-marking system of their language. In our study, this is evidenced by toddlers' failure to recognize words more efficiently when they viewed objects referring to words with different rather than the same gender. As we took special effort to ensure that the target nouns were familiar to our age groups, the determiner–noun combinations in our study should be one of the first children acquire. Had the toddlers in our study been sensitive to grammatical gender, they should have recognized target words more efficiently when determiner gender was informative, similar to the children aged 2;4 tested by Johnson (2005b). In summary, our results thus show that although Dutch two-year-olds have not yet completely mastered the determiner system of their language, they nonetheless behave similarly to their English age-matched counterparts.

While we find no evidence that Dutch learners are delayed compared to English learners in their use of articles to recognize nouns, our results do suggest that Dutch learners may be slower than French learners in their utilization of gender information on articles to facilitate word recognition. By age 2;1, French-learning toddlers tested with the Preferential Looking Paradigm recognize words more efficiently when they are presented with two objects referring to nouns with different rather than same gender (e.g. Van Heugten & Shi, 2009; also see Lew-Williams & Fernald, 2007, for data with older Spanish-learning children). Interestingly, production studies corroborate these results. Dutch learners are delayed (e.g. Rozendaal & Baker, 2008) and persist in making gender-marking errors long after French-learning children (Van der Velde, 2003). Several possible explanations may account for this apparent delay in article mastery. First, articles are more obligatory in French than Dutch (as discussed in Rozendaal & Baker, 2008). Second, definite articles are more reduced in Dutch than in French, which may make articles more salient in French. Third, noun gender has been argued to be more phonologically transparent in French than in Dutch (Corbett, 1991), which may speed the acquisition of these items in French relative to Dutch. Fourth, gender agreement is evidenced more in French than in Dutch. Dutch marks gender on its definite articles and demonstratives, but not on indefinite articles. French, in contrast, does make a difference between masculine and feminine indefinite articles (i.e. *un* vs. *une*). In addition, gender is more pronounced with regard to adjective agreement in French than in Dutch. That is, unlike French, gender distinction on the adjective is only marked after indefinite articles in Dutch

(e.g. *een mooie bal* 'a nice ball' vs. *een mooi boek* 'a nice book' but *de mooie bal* 'the nice ball' vs. *het mooie boek* 'the nice book'). Lastly, the notion that *het* can be used in other syntactic functions (i.e. expletive pronoun, provisional subject) could complicate article acquisition for Dutch toddlers. It is likely that a complex combination of all of these factors delays determiner acquisition in Dutch compared to French. Thus, our comparison of Dutch and French serves as a prime example of how function word use in comprehension could be impacted by a myriad of language-specific factors.

In future work, it would be interesting to compare the use of articles in speech comprehension in languages varying along other dimensions such as semantic transparency or number of definite articles. German, for example, would make a good comparison as it is structurally very close to Dutch (and English) but, unlike Dutch, its articles are both gender and case marked, leading to a higher variability in surface form. While there is evidence that German learners use definite articles to categorize novel words at an early age (Höhle, Weissenborn, Kiefer, Schulz & Schmitz, 2004), no study to date has looked at the use of determiners for noun comprehension.

The effect that toddlers performed better in correct than nonsense trials held regardless of whether the target was a *de* or *het* word. At this stage, children thus accept both article forms to precede nouns of either gender, implying that both forms are acquired roughly at the same rate. Production studies, however, show an asymmetrical acquisition pattern, with *de* being acquired earlier than *het* (e.g. Van der Velde, 2003). A possible explanation for this asymmetry can be found in the fact that *de* only serves as a definite article, while *het* can take on other grammatical functions. This increases the complexity of *het* relative to *de* and also drives down the frequency of *het* as a definite article. Although the number of occurrences of *de* and *het* in our corpus count is nearly identical, the number of occurrences where *het* is used as a definite article is considerably lower.[5] Dutch toddlers may thus be expected to learn the more frequent and less complex article *de* first.

Why did we fail to observe an asymmetry in the use of *de* and *het* in our study? Perhaps the most straightforward possibility is that we would have found evidence for this hypothesis had we tested younger children. Alternatively, if children initially do not distinguish between *het* as a definite article and *het* as a pronoun (see Shi & Lepage, 2008, for support that infants may initially not differentiate between function morpheme homonyms), then at the outset *de* and *het* would be equally frequent in the input. Asymmetries between *de* and *het* only appear during later stages of

[5] The Van de Weijer corpus, a corpus in which words are tagged for their linguistic category, shows that *het* is used as a pronoun more than as an article (*het* is an article 48% of the time in child-directed speech, 25% in infant-directed speech and 33% in adult-directed speech).

development when children start analyzing *het* in a more mature fashion (i.e. when they realize that it can have different syntactic functions). However, this later differentiation may not be robust enough to trigger an asymmetry effect in this study. The multiple syntactic uses of *het* may thus complicate article acquisition, but only at later stages.

Regardless of the question why *de* and *het* behave similarly, this study shows that infants are sensitive to the systematic co-occurrence of both *de* and *het* with nouns before they know which article precedes which specific word. The development of determiners thus follows a stepwise pattern, with more general preceding more detailed article knowledge, an effect that could only be found by examining the acquisition pattern in a gender-marking language.

REFERENCES

Brown, R. (1973). *A first language*. Cambridge, MA: Harvard University Press.

Christophe, A., Millotte, S., Bernal, S. & Lidz, J. (2008). Bootstrapping lexical and syntactic acquisition. *Language and Speech* **51**, 61–75.

Clark, E. V. (1993). *The lexicon in acquisition*. Cambridge: Cambridge University Press.

Corbett, G. (1991) *Gender*. Cambridge: Cambridge University Press.

Demuth, K., Culbertson, J. & Alter, J. (2006). Word-minimality, epenthesis, and coda licensing in the acquisition of English. *Language and Speech* **49**, 137–74.

Echols, C. H. & Newport, E. L. (1992). The role of stress and position in determining first words. *Language Acquisition* **2**, 189–220.

Gerken, L. A., Landau, B. & Remez, R. E. (1990). Function morphemes in young children's speech perception and production. *Developmental Psychology* **26**, 204–216.

Gerken, L. A. & McIntosh, B. (1993). The interplay of function morphemes and prosody in early language. *Developmental Psychology* **29**, 448–57.

Gleitman, L. R. & Wanner, E. (1982). Language acquisition: The state of the art. In E. Wanner & L. R. Gleitman (eds), *Language acquisition: The state of the art*, 3–48. Cambridge: Cambridge University Press.

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**, 341–53.

Johnson, E. K. (2005a). English-learning infants' representations of word forms with iambic stress. *Infancy* **7**, 99–109.

Johnson, E. K. (2005b). 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*, 320–30. Sommerville, MA: Cascadilla Press.

Johnson, E. K. & Diks, M. (2005). On-line processing of grammatical gender in Dutch-learning toddlers. Paper presented at the Xth International Congress for the Study of Child Language, Berlin, Germany.

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**, 325–38.

Lew-Williams, C. & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science* **18**, 193–98.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*, 3rd edn. Mahwah, NJ: Lawrence Erlbaum Associates.

Morgan, J. L., Shi, R. & Allopenna, P. (1996). Perceptual bases of rudimentary grammatical categories: Toward a broader conceptualization of bootstrapping. In J. L. Morgan &

K. Demuth (eds), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 263–83. Mahwah, NJ: Lawrence Erlbaum Associates.
Rozendaal, M. I. & Baker, A. E. (2008). A cross-linguistic investigation of the acquisition of the pragmatics of indefinite and definite reference in two-year-olds. *Journal of Child Language* **35**, 773–807.
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.
Shipley, E., Smith, C. & Gleitman, L. (1969). A study in the acquisition of language. *Language* **45**, 322–42.
Taelman, H., Durieux, G. & Gillis, S. (2009). Fillers as signs of distributional learning. *Journal of Child Language* **36**, 323–53.
Van der Velde, M. (2003). Déterminants et pronoms en néerlandais et en français: syntaxe et acquisition. Unpublished PhD dissertation. Université de Paris 8, Paris.
Van Heugten, M. & Shi, R. (2009). French-learning toddlers process gender information on determiners in word comprehension. *Developmental Science* **12**, 419–25.
Wijnen, F. & Verrips, M. (1998). The acquisition of Dutch syntax. In S. Gillis & A. De Houwer (eds), *The acquisition of Dutch*, 223–99. Amsterdam: Benjamins.
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**, 199–231.