The impact of lexical-statistical learning on Dutch children's use of fillers and articles

Helena Taelman & Steven Gillis

CNTS, University of Antwerp

Address for correspondence:

Helena Taelman,

Land Van Waaslaan 187,

B-9040 Gent,

Belgium

Helena.Taelman@telenet.be

At the time of this study the first author was postdoctoral researcher for the National Foundation of Science - Flanders. This study was supported by a BOF grant awarded by the University of Antwerp to the first author. We would like to thank Inge Molemans for the annotation of the data.

The impact of lexical-statistical learning on Dutch children's use of fillers and articles

Abstract

Young children sometimes insert seemingly meaningless filler syllables in the position of articles. This study reports on such fillers in ten Dutch-speaking children between age 1;0 and 3;0. Their usage reflects children's immature understanding of the articles' distribution, which is highly determined by their lexical-statistical features: most fillers occur after lexical items that frequently precede articles in the input. These words act as 'anchor' words that attract fillers in the children's speech. Such anchor words also affect the acquisition of full-fledged articles.

INTRODUCTION

Before the age of three, some children scatter FILLERS over their utterances, i.e. seemingly meaningless syllables that do not correspond to conventional words. The utterances in (1) exemplify the forms that these fillers can take: the English child in (1a) and the French child in (1b) use single vowels or nasals, the German child in (1c) produces a consonant vowel sequence, and the English child in (1d) generates a string of meaningless syllables.

(1a) [m] pick [ə] flowers	(Seth, 1;10, from Peters & Menn, 1993)
(b) [a] camion [e] Sophie	(Sophie, 1;11 from Kilani-Schoch & Dressler, 2001)
(c) [hə] weint	(Bernd, 2;7, from Vollmann, 1997)
(d) [nɪnɪnɪ] the bagel	(Steven, 2;1, from Feldman & Menn, 2003)

These fillers often serve as proto-grammatical devices (see Peters, 2001a for a discussion of other functions). Children use them instead of function words that they have not mastered yet: the fillers then appear in positions where function words are expected to occur, and/or they reflect the phonological shape of the targeted function word. But the distributional and phonological correspondence with the target is not perfect yet due to the child's immature understanding of the adult distribution and phonological shape. Sometimes, fillers mark a grammatical slot that can be filled with various alternatives, which the child still needs to disentangle. For instance, the English boy Seth produced nasal fillers in front of lexical verbs, in contexts where adults would supply modal verbs such as 'want', 'gonna', 'let', or 'can' (Peters, 2001b). These nasal fillers functioned as 'proto-modals', which signaled the child's awareness that some modal element should fill the slot, while he was still unable to differentiate between the various modal verbs. Other studies observed fillers functioning as proto-question words (Dabrowska, 2001) or proto-articles.

This last function is very common. Fillers act as proto-articles in French-, German-, English-, Spanish- and Dutch-speaking children (Bassano & Eme, 2001; Kilani-Schoch & Dressler, 2001; Lleo, 2001; Lopez-Ornat, 2001; Pepinsky, Demuth & Roark, 2001; Peters & Menn, 1993; Taelman, Durieux & Gillis, 2008; Vollmann, 1997). These proto-articles reflect an immature understanding of the shape and distribution of articles in the target language: their phonological shapes are often approximations of the target forms and they typically (but not exclusively) appear at the beginnings of NPs, though with a lower frequency than determiners in the target language. During the period in which children rely on proto-articles, many NPs still have an empty determiner slot (Lopez-Ornat, 2001).

Which factors drive the in/exclusion of a proto-article? Studies of French-speaking children suggest an important role for rhythmic preferences. Before the age of two, French children prefer to insert proto-articles in front of monosyllabic nouns in order to achieve an iambic rhythm of a weak syllable followed by a strong one, which is a dominant rhythm in the input (Bassano, Maillochon & Mottet, 2008; Demuth & Tremblay, 2008; Veneziano & Sinclair, 2000). A case study of an English-speaking child reported on a rather different rhythmic effect, related to the prosodic status of the noun's initial syllable: proto-articles were less frequent when the noun's initial syllable was weak (Pepinsky, Demuth & Roark, 2001). This effect was probably caused by a general rhythmic constraint against two consecutive weak syllables, which has been reported to affect English children's productions of multisyllabic words and grammatical morphemes as well (e.g. Gerken, 1996). We found the same tendency in a case study of a Dutch two-year-old girl (Taelman, Durieux & Gillis, 2008).

Unlike the French and the English studies, the rhythmic factor was not the main organizing principle underlying the use of the proto-articles in our Dutch case study. From the very start, their occurrence was highly determined by the preceding lexical context. They were particularly frequent after the words *ook* ('also'), *nog* ('another')[i], *is* ('is'), *gaat* ('goes'), and *niet* ('not'): between age 1;9 and 2;3 more than 70% of all proto-articles followed one of these 'anchor' words (as we will call them henceforth). Moreover, after an anchor word, the child was much more likely to supply a proto-article in determiner position as in (2a), than in any other context: while the frequency of proto-articles at the beginnings of NPs on the total number of NPs was less than 13% in normal circumstances, it increased to 78% if an anchor word preceded (between 1;9 and 2;3). Anchor words even attracted proto-articles when no NP followed, as in (2b), what happened in 59% of all tokens that did not precede an NP.

(2a)	dit is ə paard	(Cato, 2;0)
	this is F horse	
(b)	tijger is ə wakker	(Cato, 2;4)
	tiger is F awake	

An analysis of child directed speech revealed that these highly frequent sequences of anchor words and proto-articles reflected highly frequent sequences of words and articles in the input, indicating that the child had picked up these patterns from the input through statistical learning of word adjacencies: the child preferred to produce proto-articles after words that were frequently followed by articles in the input. Apparently, the child's immature understanding of the distribution of articles was strongly determined by these lexical-statistical tendencies.

This observation nicely fits with the recent attention for statistical learning as a language acquisition strategy (e.g. Saffran, Newport & Aslin, 1996). A number of studies reported that morphosyntactic acquisition is often influenced by a lexical-statistical learning strategy: function words and inflections emerge earlier in some word contexts than in others ('low-scope' or 'item-based' learning, e.g. Tomasello, 2000), and the choice of these word contexts appears to be influenced by statistical regularities in the input (e.g. Theakston, Lieven, Pine & Rowland, 2002). Our case study showed that this strategy not only affects function words, but also their precursors, i.e. fillers with a proto-grammatical function.

But how common is this lexical-statistical learning process? For how long do children rely on it, and what is its impact on the construction of their grammatical knowledge? These questions are burning issues in the field of morphosyntactic acquisition, and they certainly apply to the phenomenon of fillers acting as proto-articles. So far, our case study is the only study on fillers that attests an impact of lexical-statistical learning, which may suggest that it is an optional and uncommon learning technique for the production of fillers. But is this really the case? If children rely on statistical learning when using fillers, does it also affect the use of full-fledged articles? What is its impact on children's grammatical understanding of articles? Do the highly frequent sequences of words and articles in the input help them to derive the grammatical position of (proto-)articles, i.e. at the beginnings of NPs? Do they help them to infer that (proto-)articles are not entirely optional, but that Dutch NPs often require a filled determiner slot? One might think of a scenario where the anchor words guide the entire learning process, from the emergence of (proto-)articles over their association with NPs to the identification of their obligatory contexts.

We know already that children do not master these grammatical features instantaneously, at least not in production. Although they discover articles in the input quite early and appear to have accurate phonological representations of articles around 1;1 (Shi, Werker & Cutler, 2006), they tend to omit articles from their early utterances, or replace them by proto-articles, and need quite some time to gain an adultlike level of productivity. Dutch children are particularly slow in this process, as compared to English and French children (Rozendaal & Baker, in press; Van der Velde, 2003): at 2;6, articles and other determiners only occur in 50% of the obligatory contexts, while French children of the same age supply a determiner in 80% of the NPs (Bassano, Maillochon, Mottet, 2007) and English children surpass the 50% level already by two years of age (Abu-Akel & Bailey, 2000). In spite of the high omission rates, several studies claim that children display knowledge of the appropriate grammatical position of articles from early on: when they produce them, they do so in the correct contexts (e.g. Valian, 1986; Bottari, Cipriani & Chilosi, 1993/1994; Bohnacker, 1997). However, two studies on the acquisition of articles by English-speaking two-year-olds attested a few errors in their data: articles occurred outside NPs, in front of proper nouns, or after another determiner (Pine & Lieven, 1997; Pine & Martindale, 1996). Such errors also occurred in our Dutch case study.

The English studies by Pine et al. are the only ones to attest an impact of lexical-statistical learning in children's production of articles. They found that English-speaking children often produced articles in contexts such as *there's a + X*, *want a + X*, which frequently occur in the ambient speech. A few such lexical contexts accounted for more than half of the realizations of determiners (mostly articles). Unfortunately, we do not know whether this lexical factor influenced full-fledged articles or proto-articles, or both, since these categories were not distinguished explicitly in their study.[ii] On the basis of contrastive analyses of proto-articles and full-fledged articles, we concluded in our previous case study that the influence of lexical-statistical learning was limited to the child's proto-articles, and did not extend to the child's realizations of full-fledged articles. This conclusion was based on the fact that full-fledged articles were not more frequent after the anchor words that attracted proto-articles. However, we did not pursue the possibility that full-fledged articles in the child's speech may have been tied to other anchor words than the proto-articles were. In other words, the set of anchor words that attracted proto-articles could have been different from the set of words that attracted full-fledged articles. As a result, the evidence concerning the scope of the lexical-statistical learning strategy is still inconclusive.

The aim of the current investigation is to replicate our previous observation about the impact of lexical-statistical learning on the production of fillers for a larger group of Dutch children. In adddition, we will conduct a more refined analysis of the impact of this learning strategy on children's realizations of full-fledged articles in order to determine the scope of this strategy. Finally, we will discuss the relationship between children's sensitivity to lexical context and their growing grammatical knowledge of articles. To this end, we will analyze naturalistic, longitudinal data of ten Dutch-speaking children from the CLPF database (Fikkert, 1994; Levelt, 1994; available through CHILDES, MacWhinney, 2000). This database was selected because of its narrow phonetic transcription, which is a prerequisite for distinguishing fillers from full-fledged articles, Moreover, the children are between 1;0 and 3;0, which is exactly the age range for the acquisition of articles. The data of the child in our previous study are part of the same database, and will be included to facilitate comparison.

METHOD

Data

The Dutch CLPF database (Fikkert, 1994; Levelt, 1994; available through CHILDES, MacWhinney, 2000) was set up for phonological analysis. The children were recorded twice a month during play sessions of approximately 30-45 minutes. All child utterances were transcribed orthographically and phonetically by two transcribers, who sought consensus in cases of discrepant transcriptions. Adult utterances were not transcribed. The transcripts do not include contextual information.

Originally, the corpus contained data of 12 children, but we removed two of them because of data sparsity (≤1500 word tokens available). The remaining ten children's ages varied between 1;0 and 1;11 at the start of observations, and 2;3 and 2;8 at the end. The MLUw, Mean Length of Utterances counted in words, reached values between 1 and 4.4. Table 1 indicates the exact ages and MLUw values at the start and the end of observations for each child, as well as the number of sessions and the overall number of word tokens. The first child, Cato, was the subject of our previous case study.

Insert Table 1 about here

In order to facilitate comparison between children, we decided to split each child's corpus into six stages of morphosyntactic proficiency, defined in terms of

MLUw: (1) MLUw≥1, (2) MLUw≥1.3, (3) MLUw≥1.5, (4) MLUw≥2.0, (5) MLUw≥2.5, (6) MLUw≥3. From the third stage on, each stage represents a progress of 0.5. Between MLUw≥1 and MLUw≥1.5, an extra step was added at MLUw≥1.3 to reflect children's slower evolution. The end point of each stage was the last session before the child for the first time reached an MLUw value of the next stage. None of the children exceeded MLUw≥3.5 during the time of observation, except Enzo in his last session. This was excluded from analysis. Table 2 shows the available stages per child, indicates at which ages the children transferred to each stage, and how many sessions were included in each stage.

Insert Table 2 about here

The MLUw values were computed on the basis of all utterances in each session (by means of CLAN, MacWhinney, 2000). In order to make these measures independent of the phenomena under investigation, we did not count fillers and articles in the children's utterances.

Identification of fillers

In the original transcripts, fillers were not identified as such. Instead targetless vowels and nasals were often attached to the preceding or the following word on the phonetic tier, as in (3a) in which the final schwa of ['nɔkə] is actually targetless. If these sounds appeared in positions where a function word was expected, they were transcribed as function words, as the [n] in (4a) which was orthographically transcribed as the preposition *op* ('on').

(3a) nog een poes (Cato, 1;10)

/'nɔχ ən 'pus/

['nɔkə əm 'pys]

another a cat

(b) nog F een poes

/'nɔχ - ən 'pus/

['nɔk ə əm 'pys]

another F a cat

(4a) op tafel (Cato, 1;11)

/'ɔp 'tafəl/

[n 'tafɔ]

on table

(b) F tafel

/- 'tafəl/

[n 'tafɔ]

F table

For the purpose of this study, we identified fillers in the phonetic transcriptions in the following way. If the phonetic transcription of a word ended or started with an extra targetless vowel or nasal, we separated it and coded it as a filler. For instance, the form ['nɔkə] in (3a) was split into [nɔk] and [ə], and transcribed as a sequence of *nog*

(/’nɔχ/, ‘another’) plus a filler (F) as in (3b). We refrained from this interpretation if the extra vowel possibly stood for a common suffix or prefix, such as the adjectival declension suffix /ə/. If the child attached a schwa after an adjective, we did not interpret this vowel as a filler, but as an attempt at the adjectival declension suffix.

Many fillers were interpreted as function words in the original transcripts. We judged the probability of this interpretation by comparing the phonological form of the child’s utterance to the phonological form of the target word. If the child’s form did not resemble or approximate the target form closely, as in (4), we annotated it as a filler. Distinguishing articles from fillers was less evident, because of their strong resemblance to schwa fillers, i.e. fillers with the form [ə]: all Dutch articles are monosyllabic and contain a schwa (*een* /ən/ ‘a’, *de* /də/ ‘the’, *het* /(h)ət/ ‘the’). Moreover, some pronunciation variants of the indefinite article *een* (‘a’) in casual conversational Dutch, [n] and [ə], are identical to the typical form of fillers in child speech.[iii] If we would apply a rather loose criterion of phonological resemblance for these forms, [n] or [ə] would almost always be interpreted as an article. Instead, we decided to categorize these ambiguous forms as fillers, and only consider the article interpretation if it resembled its canonical pronunciation closely enough, where ‘closely enough’ was defined as follows: (1) if the phonological form contained a vowel followed by a dental stop and preceded by an optional [h], it was categorized as an attempt at the definite article *het* (‘the’), (2) *de* (‘the’) required a phonological form composed of a dental stop plus a vowel, and (3) *een* (‘a’) required a vowel plus a nasal. All other forms were annotated as fillers.

In order to determine the reliability of this annotation, 10% of all utterances in each child corpus were processed by a second, independent annotator (*N*=1834). The two annotators reached agreement in 98% of all utterances.

Analysis of NPs

As a final step, we identified all NPs in the child's speech, and tagged them as to whether a determiner would be mandatory or merely allowed according to the rules of Dutch syntax. These rules are similar, but not identical to the English rules (Haeseryn, Romijn, Geerts, de Rooij & van den Toorn, 1997). For instance, proper nouns and substance nouns (e.g. *melk*, 'milk') generally do not allow a determiner in adult Dutch. Common count nouns must be preceded by a determiner, except when pluralized and with indefinite reference, or when they occur in particular prepositional phrases (e.g. *in bad*, 'in bath'), and verbal constructions (e.g. *gitaar spelen*, 'guitar play-INF'). Since it is hard to infer whether a NP is definite or not, especially in young children's speech, we never judged a plural noun (with plural marking) to lack an obligatory determiner.

The reliability of this analysis was assessed using the same material as in the reliability analysis of the filler annotation. The annotators reached complete agreement in 97% of all utterances.

RESULTS

Although the children were not selected on the basis of their use of fillers, they all produced them. Four types of fillers were attested in the data: (1) schwas, (2) other vowels, (3) nasals, (4) sequences of a consonant and a vowel (CV). Table 3 indicates the frequency relative to the total amount of word tokens in each child corpus. The absolute number of instances of each filler type is provided between brackets. Schwa fillers are the most common type in all children, but their frequency varies considerably from 0.6% in Jarmo's data to 8.3% in Cato's. Fillers with a full vowel never have a frequency above 1%. The nasal fillers exceed this level in three children (Cato, Robin and Tirza), but have negligible quantitities in the others. Also CV fillers are infrequent overall.

Insert Table 3 about here.

In our further analyses we will focus on the schwa fillers. Since Dutch children often apply a phonological process of turning schwas into full vowels (Levelt, 2002), we will add fillers consisting of another vowel to this category without further distinction. We excluded the other categories of fillers because of their infrequency or their different behavior: the nasal fillers were phonemically conditioned to a high extent. As a first step, we will investigate whether schwa fillers function as proto-articles. Then, we will check for the existence of anchor words that attract schwa fillers, and examine their impact on children's grammatical understanding of articles. Finally, we will study the emergence of full-fledged articles in relationship to their lexical context.

Schwa fillers as proto-articles

The schwa fillers are phonologically very similar to the Dutch articles *een, de*, and *het*, as if the children inferred this form from the articles' common phonological characteristics: they are all monosyllabic, unstressed and contain a schwa. But do they share the distribution of articles? If the distribution of schwa fillers is modelled after the distribution of articles in the input, their frequency will be high in the canonical position of articles, i.e. at the beginnings of NPs, and infrequent in other positions. The last condition is crucial, since schwa fillers may reach high frequencies in both positions, for instance when children use them as a kind of glue between words. Hence, we computed the frequency of schwa fillers at the beginnings of NPs on the total number of NPs, and compared this figure with the frequency of schwa fillers in other positions on the total number of words outside NPs. These frequencies are plotted in Figures (1a-j), together with indications of the significance in a 2x2 chisquared test.

Insert Figures 1a-j about here

The interpretation of schwa fillers as (proto-)articles is supported by distributional evidence in eight out of the ten children. One child, Enzo (Figure 1d), already prefers fillers in NP-initial positions from the first MLUw stage for which data are available (MLUw≥1.5: $\chi^2(1, N=1007)=13.6$ $p<.001$). The other seven children (Cato 1a, David 1b, Elke 1c, Leon 1f, Robin 1h, Tirza 1i, Tom 1j) develop from a more or less random distribution to a preference for NP-initial positions somewhere between MLUw≥1.3 and MLUw≥2.5. Two children, Jarmo (1e) and Noortje (1g), do not display any evidence for considering schwa fillers as proto-articles. These children produce few schwa fillers anyway, and their frequency is never higher at the beginnings of NPs. We cannot exclude the possibility that they developed a preference for fillers in NPs after the observation period.

Apart from their shape, the schwa fillers differ from real articles in two ways. (1) The preference for NP-initial position is often not absolute: three children, Tirza, Cato and Tom, still produce fillers outside NPs with a frequency of up to 5-7% when they have already started to prefer NP-initial positions. (2) The fillers in NP-initial positions never reach the frequency level of articles in child-directed Dutch, when

36% of all nouns are preceded by an article.[iv] The highest frequency across children is 27.5% in Tom at MLUw≥2. The lowest frequency that is still significantly different from the frequency outside NPs, can be found in Elke at stage MLUw≥1.5, and only amounts to 3.8%. In five children (Cato, David, Enzo, Leon, Tirza,), there is a decline in the frequency of fillers towards the end of the observations. This decline was to be expected on the basis of the fact that fillers are non-adultlike elements and hence bound to disappear (Dressler & Kilani-Schoch, 2001). The other children probably start to reduce their filler productions after the end of the observation sessions.

Anchor words attracting schwa fillers

We will now investigate whether the occurrence of schwa fillers is influenced by the preceding lexical context, as in our previous case study. This analysis is only feasible if the data contain a minimum number of utterance-internal schwa fillers. We put the threshold at 50 per child. Since Elke's, Jarmo's and Tom's data do not fulfill this condition, they were excluded from the analysis. For all other children, we investigated the existence of anchor words that attract schwa fillers. Our initial criterium was their frequency in front of utterance-internal tokens of schwa fillers. We selected all words that precede more than 5% of the utterance-internal schwa fillers (over time). For instance, the word *ook* ('also') in Robin's data precedes 17% of Robin's utterance-internal tokens of schwa fillers, whereas most other words only reach frequencies of 1% before utterance-internal schwa fillers. Table 4 lists three to six anchor words per child.

Most anchor words are common function words that we expect to be frequent in child speech anyway. Can their frequency before schwa fillers be traced back to the their overall frequency, or do these anchor words really attract schwa fillers? Our data support the latter hypothesis. We determined per child per word whether the frequency of fillers after that word on the total number of tokens of that word was higher than the frequency of fillers after other words on the total number of tokens of other words (excluding the other anchor words from our list). The frequency of fillers was always higher after anchor words than after other words, which means that these anchor words really attract fillers.

Insert Table 4 about here

Unfortunately, we could not support this item-by-item analysis by means of a statistical test since too many expected frequencies in the chisquared analysis were below 5. Instead, we grouped the anchor words in one category, and compared the overall frequency of schwa fillers after these anchor words with the frequency of schwa fillers after other words by means of a chisquared test. The frequencies in the two conditions and the outcome of the chisquared test are displayed in Table 4. The difference in frequency between the conditions ranges from 10.1% (in Enzo) to 50.3% (in Cato), and is always significant.

Are these anchor words derived from the lexical-statistical characteristics of articles in the input? We will now determine their frequency before articles in adult speech. Since this is not transcribed in our database, we will rely on a large sample of Dutch child directed speech from four CHILDES databases: the Groningen database (Wijnen & Bol, 1993); Van Kampen database (Van Kampen, 1994), Wijnen database (Wijnen, 1988), and the section for normally developing children in the Bol & Kuiken database (Bol & Kuiken, 1990). The aggregated data contain 270960 utterances by adults and include 79262 realizations of *een, de,* and *het,* which were submitted to a frequency analysis (disregarding the grammatical status of these words). The articles are preceded by 1843 different words, some of which occur far more frequently than others. For instance, *nog* ('another') precedes 3.9% of all utterance-internal realizations of articles, while *beneden* ('downstairs') only precedes 0.3% of them. The most frequent words are listed in Appendix A. The ranking of each word in the frequency list is indicated as well as its frequency before articles. From the ten most frequent words in this list, six are used as anchor words by the children: *is* ('is'), *in* ('in'), *op* ('on'), *nog* ('another'), *ook* ('also'), *en* ('and'). The three other anchor words *gaat* ('goes'), *zijn* ('be') and *een* ('a'), do not have the highest frequencies in front of articles in the input, but still belong to the subtop, yielding rankings from 22 to 71 (out of 1843 words). This confirms our former hypothesis that the anchor words stem from highly frequent precedents of articles in the input.

Following the same reasoning, we may hypothesize that the children also generate more schwa fillers *before* some words over others, reflecting highly frequent sequences of articles and words (probably mostly nouns) in the input. But, applying the same methodology as in our analysis of the preceding lexical context, we could not identify such following anchor words. Almost no words passed the frequency criterium that they follow more than 5% of all instances of fillers. In one child (Noortje), we could identify one anchor word after fillers: *niet* ('not'), which appeared after 14% of all schwa fillers. 13% of all tokens of *niet* were preceded by a schwa filler (9/67), which is more than the frequency of schwa fillers before other words (2%, 61/2602; $\chi^2(1, N=2669)=31.4$ $p<.001$). The near absence of following anchor words in the children's data reflects the higher variability of words (mostly nouns) after tokens of *een, de* and *het* in the input. Our sample of child directed speech contains many more different words that follow tokens of *een, de* and *het* than words that precede tokens of *een, de* and *het*: 7149 versus 1843. As a result, the frequencies of the individual words are low as a rule.[v]

Impact on children's grammatical understanding of articles

Following the scenario proposed in the introduction, anchor words have a maximal impact on the acquisition of articles: they precipitate the emergence of (proto-)articles in children's speech, enhance their awareness of the link with NPs and help them to discover the semi-obligatory character of Dutch articles in contexts of anchor word + NP first, before this knowledge is transferred to other NPs. Does the evolution of schwa fillers conform to this scenario?

We investigated the impact of anchor words on the emergence of schwa fillers by analyzing the schwa fillers in the first MLUw stage with five or more tokens of schwa fillers. If children form the category of schwa fillers on the basis of lexical sequences with anchor words and articles in the input, we expect the majority of their first tokens of schwa fillers to occur after anchor words. But this prediction is not borne out by the data. Only two children produce their first tokens mainly after anchor words: Cato produces 14 out of 18 tokens after anchor words at MLUw≥1.3, David produces 4 out of 6 tokens after anchor words at MLUw≥1. Four other children prefer other contexts for their first tokens of schwa fillers (Enzo at MLUw≥1.5: 7/26 tokens after anchor words; Leon at MLUw≥1.3: 6/16; Noortje at MLUw≥1.3: 3/7; Robin at MLUw≥1: 1/6). One child, Tirza, produces half of his first tokens after anchor words at MLUw≥1 (5/10). Hence, we have no convincing evidence that the first realizations of schwa fillers always emerge from contexts with anchor words.

We will now analyze children's awareness of the link with NPs and their understanding of the semi-obligatory nature by means of Figures 2a-g, which display the development of fillers after anchor words in relation to their grammatical position (at the beginnings of NPs or outside NPs), and the development of fillers that follow another word, again in relation to their grammatical position. For the ease of discussion we will refer to these contexts by means of abbreviations:

- ANCHOR-NP: after an anchor word at the start of an NP, e.g. *ook _ boek* ('also book')
- ANCHOR-OUTNP: after an anchor word outside an NP, e.g. *ook _ spelen* ('also play-INF')
- OTHER-NP: after another word at the start of an NP, e.g. *weg _ boek* ('away book')
- OTHER-OUTNP: after another word outside an NP, e.g. *opnieuw _ spelen* ('again play-INF')

For each context, the frequency of fillers was computed as the number of fillers in that context over the number of times that that particular context is produced by the child. If a particular context appears less than 15 times, this data point was omitted from the figure. This often happened in the earlier stages.

Let us first address children's understanding of the semi-obligatory nature of articles. This should be expressed in high frequencies of schwa fillers at the start of NPs. Do children reach such high frequency rates more easily after anchor words than in other contexts? The answer is positive. From the moment that ANCHOR-NP constructions emerge, they immediately attract high frequencies of fillers, which differ substantially from their frequency in OTHER-NP constructions. With frequencies around 85% in ANCHOR-NP contexts, Cato (2a) is the most pronounced case. But also David (2b) and Tirza (2g) reach very high frequencies above 80%. The other children do not reach such high levels in ANCHOR-NP contexts, but the percentages exceed the percentages in OTHER-NP contexts considerably. Towards the end of observations, almost all children produce less fillers in ANCHOR-NP contexts, conforming to the general evolution of fillers. As a result, the difference between ANCHOR-NP and OTHER-NP becomes smaller and sometimes insignificant (Noortje MLUw≥2: $\chi^2(1, N=119)=2.6$ p=.110; Tirza MLUw≥2: $\chi^2(1, N=57)=3.7$ p=.056; Tirza MLUw≥2.5: $\chi^2(1, N=111)=1.5$ p=.227; Enzo MLUw≥2.5: $\chi^2(1, N=167)=2.8$ p=.096; Enzo MLUw≥3: Yates' $\chi^2(1, N=294)=2.9$ p=.088).

Insert Figures 2a-g about here

Do anchor words enhance children's awareness of the link between schwa fillers and NPs? Figure 2 shows that most children display knowledge of the link with NPs during the observation period, in contexts with anchor words as well as in contexts without anchor words. Noortje is the only exception. She neither differentiates ANCHOR-NP from ANCHOR-OUTNP (MLUw≥1.5: $\chi^2(1, N=59)=3.3$ p=.068; MLUw≥2: $\chi^2(1, N=74)=2.5$ p=.113), nor OTHER-NP from OTHER-OUTNP (MLUw≥1.3: $\chi^2(1, N=95)=0.7$ p=.397; MLUw≥1.5: $\chi^2(1, N=250)=1.1$ p=.296; MLUw≥2: $\chi^2(1, N=349)=0.5$ p=.489). Apparently, she does not know the typical position of articles yet, but has nevertheless absorbed some highly frequent co-occurrence patterns in the input.

But, instead of enlarging the contrast between NP-initial positions and positions outside NPs, anchor words seduce several children into producing more fillers outside NPs than they do in OTHER-OUTNP contexts. The most extreme case is Cato. This child already prefers fillers in OTHER-NP positions above OTHER-OUTNP positions from MLUw≥1.5 on ($\chi^2(1, N=316)=41.3$ p<.001), but does not transfer this knowledge immediately to the contexts with anchor words: during MLUw≥1.5, the frequency of fillers in ANCHOR-NP contexts is as high as their frequency in ANCHOR-OUTNP contexts ($\chi^2(1, N=151)=0.9$ p=.332). From MLUw≥2 on, Cato becomes increasingly aware of the difference between ANCHOR-NP and ANCHOR-OUTNP ($\chi^2(1, N=267)=8.1$ p=.004). But this does not prevent her from producing a fair amount of schwa fillers in ANCHOR-OUTNP contexts to a degree that they surpass the frequency of schwa fillers in OTHER-OUTNP contexts ($\chi^2(1, N=308)=205$ p<.001). These lexical contexts lead her to override her awareness of the connection between articles and NPs. Further analysis reveals that the situation at MLUw≥1.5 is due to the impact of one anchor word, *is*, which seems to constitute a holophrase together with the schwa filler at that time: the schwa appears after *is* very frequently and independently of the grammatical context, in NP-initial positions (87%, *N*=89) as well as outside (84%, *N*=45). The other anchor words already attract more schwa fillers in NP-initial positions (in 83% of the cases, *N*=12) than outside (40%, *N*=5). A similar difference can be found for *is* at the later stages (MLUw≥2: 72% in NPs versus 48% in other positions, $\chi^2(1, N=116)=7.2$ p=.007).

Leon and Tirza differ from Cato in that they barely produce fillers in ANCHOR-OUTNP contexts when these contexts first occur (with a minimum frequency of 15), i.e. at MLUw≥1.5 in Leon's case, and at MLUw≥1.3 in Tirza's case. From their equally low frequency in OTHER-OUTNP contexts, it can be concluded that the children avoid fillers outside NPs in general. Both children however produce more fillers in ANCHOR-OUTNP contexts over time. Tirza generates a reasonable amount of fillers in ANCHOR-OUTNP contexts from the moment that he starts to produce fillers in ANCHOR-NP contexts. Probably, his practice with these highly frequent sequences of anchor words and fillers in grammatical contexts seduces him into producing fillers after anchor words in ungrammatical contexts, even though this runs counter to his earlier avoidance of fillers outside NPs. This tendency however has already disappeared by the next stage. In Leon's case, the frequency of fillers in ANCHOR-OUTNP contexts gradually increases to 11% at MLUw≥2.5, and then shrinks again. Again his practice with fillers in ANCHOR-NP contexts may be responsible for the increase in ANCHOR-OUTNP contexts.

Also Robin, David and Enzo produce nonnegligible amounts of fillers in ANCHOR-NP contexts. Robin does so from the first moment that he produces his set of anchor words. David displays this pattern at MLUw≥2, after a gap in the data due to which we do not know how he evolved towards it. Enzo resembles Leon and Tirza in that he starts with very few productions in ANCHOR-OUTNP contexts, which then increase to 8.6% at MLUw≥2 and finally decrease again. The difference between ANCHOR-NP and ANCHOR-OUTNP is always relatively small, and only significant at MLUw≥2.5 ($\chi^2(1, N=107)=5.2$ p=.023).

We can conclude that schwa fillers do not necessarily make their first appearance in the lexical contexts with anchor words. But from the moment that the children start to produce constructions with an anchor word followed by an NP, these attract a high rate of schwa fillers, indicative of the children's awareness that these positions must be filled with an article-like element (in most cases). This awareness of the semi-obligatory nature is less pronounced for NPs in other contexts. The other side of the coin is that the anchor words seduce several children to produce such fillers when no NP follows. Thus, while the anchor words help children to reach a high productivity of fillers, they seduce them to override their awareness of the connection with NPs. Apart from these few generalizations, we observe interindividual differences in the way that the anchor words interact with these children's grammatical knowledge. One child does not distinguish these grammatical positions anyway. Another child neglects the grammatical context after anchor words during the first stage of use, although she does take it into account outside the context of anchor words. This child starts to develop a preference for ANCHOR-NP contexts above ANCHOR-OUTNP contexts from the next stage on. Three children prefer to produce fillers at the beginnings of NPs and avoid them outside, but are teased into violating this rule in ANCHOR-OUTNP contexts after some practice with fillers after anchor words at the beginnings of NPs. The two remaining children produce a nonnegligible amount of fillers in ANCHOR-OUTNP contexts, but nevertheless prefer fillers in ANCHOR-NP positions.

The emergence of full-fledged articles

All children start to produce full-fledged articles during the time of observations, but three children only generate them in less than 3% of all NPs (Jarmo, Elke, Noortje). These children also produce very few fillers, and most NPs (≥85%) have an empty determiner position anyway. Hence, it seems that these children are still in a very early stage of article acquisition.

The emergence of full-fledged articles in the other seven children is displayed in Figures 3a-g. Each figure plots the frequency of full-fledged articles in NPs, and for the sake of comparison, the frequency of schwa fillers at the beginnings of NPs, and the percentage of NPs with a filled determiner slot (filler, article, other determiner).

Insert Figures 3a-g about here

All children produce articles and fillers concurrently. Cato, Leon and Tirza first produce more fillers than articles before the articles surpass the frequency of fillers. David, Robin and Tom only display the first part of this developmental process: they produce more fillers than articles (although in Robin's case fillers are not very frequent.) Enzo however produces more articles than fillers right from the start. We do not know what happened before the observations. None of the children produce more articles than schwa fillers before MLUw≥2.

In general, the frequency of articles is still low. David, Robin and Tom reach a maximum frequency between 9 and 12% (of all NPs). Cato, Enzo, Leon and Tirza reach maximum frequencies between 22 and 37%. Since the schwa fillers have a high frequency at the same time, and some other determiners occur, these last four children fill between 40% and 60% of the determiner slots. For the sake of comparison, we computed the percentage of determiner positions that would be filled if they had produced a determiner in all cases that require one (following the guidelines mentioned in the method section), and plotted this by means of a dotted curve on the figures. The curve always runs at least 15% above the actual level of filled determiner slots. Apparently, even the rapidly developing children do not master this aspect of determiner acquisition entirely.

Unlike schwa fillers, articles almost never occur outside NPs. Most realizations of *een, de* and *het* outside NPs fulfill the function of a pronoun, as in (5a) or they can be analyzed as false starts that are taken up later in the child's utterance as in (5b). Some are really ungrammatical as in (5c). Most children produce at most one ungrammatical instance. Exceptions are Cato with ten ungrammatical instances, Robin with two and Tirza with four.

(5a) het staat niet zo goed (Leon, 2;5.13)

it does not fit well

(b) hier is een nog ə clown (Tirza, 2;3.27)

here is a another F clown

(c) gaat een eten (Cato, 2;4.23)

goes a eat-INF

The children master most other distributional restrictions of articles, for instance they almost never produce articles in front of proper nouns (at most one exception per child). On top, they use the singular indefinite article *een* with singular nouns only. Cato is an important exception in this respect. She produces the singular indefinite article *een* 21 times in front of a plural noun (out of a total of 291 plural nouns). One example can be found in (6a). Robin and Tirza produce such instances twice during the time of observations. The other children realize at most one instance.

(6a) allebei een pyjama's aan (Cato, 2;6.6)

both a pyjamas-PL on

(b) ik een ə koekje (Robin, 2;1.25)

I a F cookie

(c) als ə de bus gaat (Leon, 2;4.1)

if F the bus goes

(d) paardje had een zo'n appel (Cato, 2;7.4)

horse-DIM had a such-a apple

(e) een een tijger (Cato, 2;5.23)

a a tiger

Cato, Robin and Tirza also make several mistakes against the rule that an article may not be combined with another determiner. Cato produces an article next to another determiner or a filler 37 times, Robin 9 times and Tirza 15 times. A few of these errors occur in David (3x), Enzo (5x), Leon (2x), and Tom (3x) as well. Some examples can be found in (6b-e).

We can conclude that errors against the distributional restrictions of articles are specific for three children only, Cato, Robin and Tirza, each of which produce articles outside NPs and in front of proper nouns, combine *een* with plural nouns, and insert articles next to another determiner. The other children master the distributional restrictions well, except for the rule that an article may not be combined with another determiner. This conforms to the English studies of Pine et al. (1996; 1997) where sequencing errors of incorrectly sequenced determiners were most common. It seems that this aspect is most difficult to acquire.

Anchor words attracting articles

By analogy with the schwa fillers, we expect that the likelihood of occurrence of full-fledged articles is influenced by the preceding lexical context. In order to investigate the impact of the identity of the preceding word we need a minimum amount of 50 realizations of full-fledged articles in utterance-internal positions, but only five children fulfill this condition: Cato, Enzo, Leon, Robin and Tirza. For each of these children, we identified the words that precede the article realizations most often, i.e. that occur in front of at least 5% of all non-initial articles. Then, we compared per child per word the frequency with which NPs start with an article after that word with the frequency with which NPs start with an article after other words. A word was recognized as an anchor word if the frequency of articles is higher after that word than after other words. Table 5 lists three to eight anchor words per child. The table provides the overall frequency of articles in NPs if they are preceded by anchor words, and the frequency of articles in NPs if another word precedes, which is 12% to 32% lower. The difference between these two conditions is always significant (cfr. Table 5).

Insert Table 5 about here

By means of the same methodology, we tried to isolate anchor words *following* articles, i.e. highly frequent words that enhance the productivity of articles before them. But no word occurred after more than 5% of all articles, rendering this analysis without results.

Can the identified anchor words be traced back to highly frequent precedents of articles in the input? In the last column of appendix A, we indicate for each of the 14 different anchor words from Table 5 their frequency in front of articles in adult speech. Seven of them belong to the ten most frequent words in front of articles in adult speech. The other seven anchor words have lower rankings, from 22 down to 120 (out of 1843 words), but still in the upper 10% of the entire list. Thus, the sequences of anchor words and articles in the children's speech appear to be derived from highly frequent sequences of words and articles in the input, just like the sequences of anchor words and proto-articles were.

Appendix A enables us to compare the anchor words before articles with the anchor words before schwa fillers. Almost all words that functioned as anchor word before schwa fillers recur as anchor word before full-fledged articles, sometimes in the same children, sometimes in others (except *zijn* and *kijk*). This strengthens our idea that both types of anchor words (before schwa fillers and full-fledged articles) are based on the same high frequency sequences of words and articles in the input.

We have no straightforward explanation for the gaps in the ranking list of Appendix A. Whether or not a highly frequent precedent of articles in the input is translated into an anchor word probably depends on multiple factors, such as the age of acquisition of that word, the constructions in which the word is used by the child, the frequency of these constructions, the frequency of the word outside contexts with articles, etc. Moreover, we need to take into account that the child's input may deviate from our general input analysis to some extent.

Let us now study the impact of the anchor words on the acquisition of articles. One might hypothesize that children's first tokens of articles happen in constructions with anchor words. This is however not the case. Per child, we analyzed the first MLUw stage with minimally five tokens of articles. These initial tokens did not occur exclusively in constructions with anchor words. Even more, four children produced most initial tokens in other contexts (Cato at MLUw≥1.5: 1 out of 36 tokens after anchor words; Enzo at MLUw≥1.5: 7/21; Robin at MLUw≥1.3: 1/5; Tirza at MLUw≥1: 0/6). Only Leon had more initial tokens after anchor words than in any other position (MLUw≥1.3: 5/8).

When does this lexical factor emerge? Figures 4a-e display the development of articles after anchor words at the beginnings of NPs, and after other words at the beginnings of NPs. From these figures it can be seen that most children (except Enzo) do not immediately produce constructions with anchor words (from the set of anchor words that attract articles). When the constructions with anchor words appear, they do not immediately attract high frequencies of articles. For instance, Leon only produces articles after anchor words with a frequency of 8% in the first stage where they emerge, while his fillers immediately yielded a frequency of 32% after anchor words. The frequency of articles after anchor words is more or less equal to the frequency of articles after other words until MLUw≥2.5 (MLUw≥1.5: $\chi^2(1, N=103)=0.0$ p=.797; MLUw≥2: $\chi^2(1, N=86)=0.2$ p=.985). Cato does not distinguish between the contexts with and without anchor words either at MLUw≥1.5 ($\chi^2(1, N=159)=1.1$ p=.292), and Tirza does not at MLUw≥2 ($\chi^2(1, N=57)=1.0$ p=.319). Since these three children produce increasingly more articles after anchor words, the difference with other contexts grows significantly over time: in Leon's case this happens at MLUw≥3 ($\chi^2(1, N=126)=6.2$ p=.013), in Cato's case from MLUw≥2 on ($\chi^2(1, N=289)=51.3$ p<.001), and in Tirza's case at MLUw≥2.5 ($\chi^2(1, N=111)=24.4$ p<.001). Enzo already produces more articles after anchor words than after other words from the start (MLUw≥1.5: $\chi^2(1, N=128)=6.0$ p=.014). Robin waits until MLUw≥2.5 before producing constructions with anchor words, and immediately produces more articles after them ($\chi^2(1, N=475)=14.8$ p<.001).

Insert Figures 4a-e about here

The initial low frequency rates after anchor words may be explained by the interplay between schwa fillers and articles. Some anchor words in some children attract articles as well as schwa fillers, for instance, the words *op* and *in* in Leon's speech. In these cases, the child probably first prefers schwa fillers after such anchor words, and then replaces them by full-fledged articles. As a result, the frequency of full-fledged articles will be low at the start of observations while the total frequency of articles and schwa fillers together is already at a high level. If this explanation is true, the low frequencies after anchor words will disappear if we include schwa fillers in our counts. This new variable is plotted in Figures 4a-e by means of a dotted curve. As predicted, the curve starts at a much higher frequency level than the curve of articles after anchor words in all children. These two curves converge after a while as a result of the declining frequency of schwa fillers (except in Robin). We can conclude that, if we include schwa fillers, the frequency of articles is high after anchor words, from the first moment that they appear.

SUMMARY AND DISCUSSION

In this study we have analyzed the impact of lexical context on ten Dutch children's productions of fillers and full-fledged articles. As a first step, we established that these children produced fillers that acted as proto-articles. The by far most frequent phonological shape of their fillers was a schwa, which incorporates the common phonological characteristics of the three Dutch articles *een, de* and *het*. In eight out of ten children, these schwa fillers were or grew more frequent in the canonical positions of articles at the beginnings of NPs than in other positions outside NPs. The two remaining children had few schwa fillers anyway. The children that produced a considerable number of schwa fillers, also realized (a considerable amount of) full-fledged articles during the observations. But the full-fledged articles never surpassed the frequency of schwa fillers before MLUw≥2, and often only after the end of the observations. These results are particularly revealing since we did not select the children on the basis of the occurrence of fillers, and hence, we may infer that the acquisition of articles in Dutch commonly includes a stage with schwa fillers acting as proto-articles.

Could the lexical context of an utterance influence children's use of (proto-)articles? We pursued whether these (proto-)articles appeared more often after some words than after others, and whether their frequency in these contexts on the total number of occurrences of these contexts was higher than their frequency after other words. Those analyses were not possible for all children, due to a lack of data: three children did not produce enough proto-articles, five children did not produce enough full-fledged articles yet. But the children who did produce enough tokens, were all influenced by the preceding lexical context. Per child we isolated a set of three to eight anchor words that frequently preceded proto-articles and/or articles and after which the frequency of proto-articles and/or articles was much higher than in other contexts. We found good evidence that these sequences of anchor words and (proto-)articles were derived from highly frequent sequences of words and articles in the input. Eight out of 17 anchor words belonged to the top ten of most frequent words before articles in the input, the other nine anchor words were still part of the 10% most frequent words in front of articles in the input. A parallel effort to discover anchor words following proto-articles or articles failed for all children but one, because there were almost no words that followed proto-articles or articles with a high frequency, which corresponded to the higher variability in lexical contexts after articles in the input.

In this way, we replicated the observation of our previous case study that children may base their early productions of proto-articles (in the form of schwa fillers) on lexical-statistical regularities of articles in the input: they produce more proto-articles after words that frequently precede articles in the input. Our study shows that the impact of this lexical-statistical learning process is not child-specific, but widespread, at least in Dutch child language. Furthermore, we also observed its influence in children's realizations of full-fledged articles, whereas our previous study had concluded that the impact of lexical context was limited to proto-articles only.

This means that lexical-statistical learning is not a temporary learning process that is given up in favor of a more mature learning principle when children start to produce full-fledged articles, but that it is a pervasive phenomenon.

Nevertheless, the existing literature on fillers and articles seems to suggest that the lexical-statistical factor is not universal. An impact of lexical context has been observed in English children's determiners (Pine & Martindale, 1996; Pine & Lieven, 1997). But French studies on fillers and articles do not attest this lexical factor, even though they are detailed and numerous (Veneziano & Sinclair, 2000; Demuth & Tremblay, 2008; Bassano, Maillochon & Mottet, 2008). It remains to be investigated whether this difference holds true in a direct comparison between these languages, and which factors are instrumental in the different status of lexical statistical-learning in French-speaking children's acquisition of articles as compared to that process in English and Dutch.

How essential was this lexical factor for children's grammatical mastery of articles? In the introduction, we proposed a scenario in which children derive the basic grammatical characteristics of articles from highly frequent sequences of words and articles in the input: they precipitate the emergence of (proto-)articles in children's speech, they help them to link (proto-)articles with the positions at the beginnings of NPs, and children would infer from them that (proto-) articles are semi-obligatory elements in contexts with anchor word + NP first.

Our data do not obey this scenario. First of all, we found that children's first realizations of schwa fillers generally happened outside the contexts with anchor words, except in two children who produced the majority of first schwa fillers after anchor words. Similarly, most initial articles did not occur after anchor words (except in one child).

Second, several children seemed to know the link between (proto-)articles and NPs from early on (except one). Instead of strengthening this link, the contexts with anchor words often caused children to override it and produce schwa fillers outside NPs. The most extreme case was Cato, who initially produced as many fillers after anchor words outside NPs as at the beginnings of NPs, while she already displayed a preference for fillers at the beginnings of NPs over other morphosyntactic positions outside the context of anchor words. We could trace back this behavior to one particular anchor word, *is*, which was almost always followed by a schwa filler irrespective of the grammatical context, as if it was a holophrase. The other anchor words also attracted several fillers when they occurred outside NPs, but less than at the beginnings of NPs. In contrast with Cato, other children initially avoided fillers after anchor words when no NP followed to the same extent as they avoided fillers in other contexts outside NPs. But after some experience with anchor word + filler sequences at the beginnings of NPs, they were seduced into producing a few fillers after anchor words outside NPs. Still other children immediately produced a nonnegligible number of fillers after anchor words outside NPs although they were less frequent than fillers after anchor words at the beginnings of NPs. There was thus some interindividual variation in the moment and the extent to which children's grammatical knowledge could be overridden by their knowledge of lexical-statistical regularities. Common to these deviations was that they concerned realizations of schwa fillers, not full-fledged articles. Apparently, children's knowledge of the connection with NPs was deeper entrenched in case of articles than in case of proto-articles.

The only grammatical feature that was acquired more easily through lexical-statistical learning, was that NPs generally require an article-like element. Indeed, most children produced very high percentages of (proto-)articles in contexts where the NP followed an anchor word. However, children's knowledge about the semi-obligatory nature of articles did not immediately (or entirely) transfer to NPs that did not follow an anchor word, which attracted a considerably lower number of (proto-)articles. They had not translated yet this context-specific knowledge into an abstract feature applicable to all NPs.

On the basis of this evidence, the lexical-statistical learning process appears to be less powerful than our scenario proposed, but its role is not insignificant either. On the one hand, children produced their first (proto-)articles in other contexts than lexical-statistical learning would predict. Furthermore, the children had already knowledge of the connection between articles and NPs when they began to produce proto-articles. It appears that this is an early accomplishment, which conforms to the results of existing comprehension studies on children's understanding of articles (Höhle, Weissenborn, Kiefer, Schulz & Schmitz, 2004). On the other hand, the lexical-statistical learning process helped these children to gain high frequencies of (proto-)articles in NPs, conforming to the semi-obligatory nature of articles. However, this process had not resulted yet in abstract grammatical knowledge applicable to all NPs.

Besides their inclination to omit articles, the children showed some problems with the finer distributional restrictions of articles. Most difficult to acquire was the rule that articles may not be combined with other function words in the category of determiners. Almost all children violated this restriction on the sequencing of determiners a few times. On top, three children also made errors against the restriction that articles may not precede proper nouns, and they sometimes combined the singular article *een* with a plural noun. The same children occasionally produced a full-fledged article in an ungrammatical position outside NPs. Also the English children in Pine et al. (1996, 1997) made such errors, particularly against the restriction on the sequencing of determiners. This aspect is probably harder to learn since it requires children to discover similarities in the distribution and function of the various determiners, and this can only proceed through the comparison of several utterances. In fact, children can only master it through the construction of an abstract determiner slot.

That our Dutch-speaking children still violated the restriction on the sequencing of determiners when they already knew that (full-fledged) articles are linked with nouns, shows that the acquisition of these two grammatical features does not necessarily proceed simultaneously. This goes against (generativist) theories, which claim that children have an adult-like functional category of determiner from early on, and that the discovery of the determiner function of articles is therefore an easy and relatively straightforward task, which includes at the same time their understanding of the link with nouns and their awareness of the restriction on the sequencing determiners (e.g. Valian, 1986). Rather we propose that the acquisition of the grammatical features of articles proceeds in three (more or less) consecutive steps. 1. Children identify the phonogical shape of articles and at the same time learn to connect articles with nouns. Since both processes are based on an analysis of the input, a better specified phonological shape goes along with a deeper entrenched link between articles and nouns. 2. They learn that articles are semi-obligatory. This happens in a context-based manner: children discover this feature in some lexical contexts faster than others. 3. They acquire some understanding of the restrictions on the sequencing of determiners. Neither the second nor the third step is entirely accomplished at the end of our observations. Future studies of children's productions as well as their comprehension will shed more light on the timetable and the driving forces of these accomplishments.

Table 1. Overview of the 10 children from the CLPF database

Child	*Ages*	*Number of sessions*	*Word tokens*	*MLUw*[1]
Cato	1;10.11-2;7.4	17	5512	1.1-3.3
David	1;11.8-2;3.25	6	2003	1.2-2.9
Elke	1;6.25-2;4.29	19	1854	1.1-1.9
Enzo	1;11.8-2;6.11	16	6023	1.9-4.4
Jarmo	1;4.18-2;4.1	23	1843	1.0-1.4
Leon	1;10.1-2;8.19	23	5186	1.2-3.2
Noortje	1;7.14-2;11.0	21	2792	1.0-2.2
Robin	1;5.11-2;4.28	23	5089	1.0-2.5
Tirza	1;7.9-2;6.12	20	3309	1.0-2.6
Tom	1;0.24-2;3.2	26	2272	1.0-2.0

[1] Mean Length of Utterances counted in Words in the first and final session (articles and fillers excluded from the counts)

Table 2. Age-MLUw correspondences. Number of sessions per stage between brackets.

Child	*MLUw≥1*	*MLUw≥1.3*	*MLUw≥1.5*	*MLUw≥2*	*MLUw≥2.5*	*MLUw≥3*
Cato	1;10.11 (3)	1;11.22 (1)	2;0.6 (5)	2;2.29 (4)	2;5.9 (3)	2;7.4 (1)
David	1;11.8 (1)			2;1.24 (1)	2;2.14 (4)	
Elke	1;6.25 (14)	2;2.6 (2)	2;3.27 (3)			
Enzo			1;11.8 (5)	2;1.3 (4)	2;3.14 (3)	2;4.25 (3)
Jarmo	1;4.18 (18)	2;1.22 (5)				
Leon	1;10.1 (2)	1;10.15 (5)	2;0.24 (7)	2;4.1 (4)	2;5.27 (3)	2;7.22 (3)
Noortje	1;7.14 (11)	2;6.19 (2)	2;7.17 (4)	2;9.2 (4)		
Robin	1;5.11 (8)	1;9.1 (4)	1;11.6 (2)	2;0.4 (1)	2;0.18 (3)	
Tirza	1;7.9 (6)	1;11.8 (6)	2;2.0 (3)	2;3.12 (2)	2;5.5 (3)	
Tom	1;0.24 (18)	1;9.4 (1)	1;10.8 (6)	2;3.2 (1)		

Table 3. Frequency of four filler types relative to the total number of word tokens

Child	*Schwa*	*Other vowel*	*Nasal*	*CV*	*All*
Cato	8.3 (*N*=458)	0.2 (*N*=10)	2.8 (*N*=152)	0.1 (*N*=3)	11.4 (*N*=623)
David	5.0 (*N*=100)	0.2 (*N*=4)	0.2 (*N*=3)	0.1 (*N*=2)	5.5 (*N*=109)
Elke	1.2 (*N*=22)	0.4 (*N*=7)	0.1 (*N*=1)	0.1 (*N*=1)	1.8 (*N*=31)
Enzo	2.2 (*N*=129)	0.2 (*N*=9)	0.4 (*N*=21)	0.1 (*N*=5)	2.9 (*N*=164)
Jarmo	0.6 (*N*=11)	0.2 (*N*=3)	0.1 (*N*=2)	0.2 (*N*=3)	1.1 (*N*=19)
Leon	3.1 (*N*=162)	0.3 (*N*=17)	0.1 (*N*=4)	0.1 (*N*=6)	3.6 (*N*=189)
Noortje	3.4 (*N*=94)	0.7 (*N*=21)	0.0 (*N*=1)	0.2 (*N*=7)	4.3 (*N*=123)
Robin	3.1 (*N*=156)	0.3 (*N*=13)	1.2 (*N*=62)	0.4 (*N*=21)	5.0 (*N*=252)
Tirza	4.8 (*N*=156)	0.8 (*N*=27)	1.4 (*N*=45)	0.8 (*N*=26)	7.8 (*N*=254)
Tom	2.0 (*N*=44)	1.0 (*N*=23)	0.2 (*N*=4)	0.2 (*N*=4)	3.4 (*N*=75)

Table 4. Words that are followed by a schwa filler more often than chance (the percentage of occurrences together with schwa on the total number of tokens)

Child	*Anchor words*	*Frequency after anchor words*	*Frequency after other words*	*Chisquared test*
Cato	is (‘is’), in (‘in’), ook (‘also’), op (‘on’), gaat (‘goes’),	55.7	5.2	$\chi^2(1, N=2641) = 855.0$; $p<.001$
David	Is (‘is’), in (‘in’), ook (‘also’), op (‘on’), nog (‘another’), en (‘and’)	31.4	4.3	$\chi^2(1, N=950)=126.6$; $p<.001$
Enzo	is (‘is’), in (‘in’), ook (‘also’)	12.3	2.2	$\chi^2(1, N=2838)=85.6$; $p<.001$
Leon	is (‘is’), in (‘in’), ook (‘also’), op (‘on’), nog (‘another’)	28.4	2.4	$\chi^2(1, N=2279)=337.8$; $p<.001$
Noortje	ook (‘also’), op (‘on’), en (‘and’), kijk (‘look’),	25.3	2.5	$\chi^2(1, N=861)=105.9$; $p<.001$
Robin	is (‘is’), in (‘in’), ook (‘also’), op (‘on’), een (‘a’)	18.6	3.2	$\chi^2(1, N=2491)=146.6$; $p<.001$
Tirza	is (‘is’), in (‘in’), nog (‘another’), een (‘a’), zijn (‘be/are’)	18.7	5.4	$\chi^2(1, N=1361)=47.5$; $p<.001$

Table 5. Words that are followed by an article more often than chance (the percentage of occurrences together with *een*/*de*/*het* on the total number of tokens)

Child	*Anchor words*	*Frequency after anchor words*	*Frequency after other words*	*Chisquared test*
Cato	ook ('also'), met ('with'), gaat ('goes'), een ('a')	47.3	15.6	χ^2(1, *N*=764)=70.6; p<.001
Enzo	op ('on'), ook ('also'), is ('is'), dit ('also')	53.4	22.1	χ^2(1, *N*=614)=55.3; p<.001
Leon	op ('on'), in ('in'), met ('with'), gaat ('goes'), van ('of')	28.9	13.8	χ^2(1, *N*=468)=14.4; p<.001
Robin	op ('on'), in ('in'), ook ('also'), met ('with'), is ('is'), niet ('not'), uit ('off'), heb ('have-1/2SG')	16.7	4.9	χ^2(1, *N*=593)=22.8; p<.001
Tirza	op ('on'), in ('in'), is ('is'), niet ('not'), nog ('another'), zit (sit-1/2/3SG)	44.9	12.5	χ^2(1, *N*=298)=38.8; p<.001

Figure 1. The frequency of schwa fillers at the beginnings of NPs and outside NPs in the consecutive MLUw stages

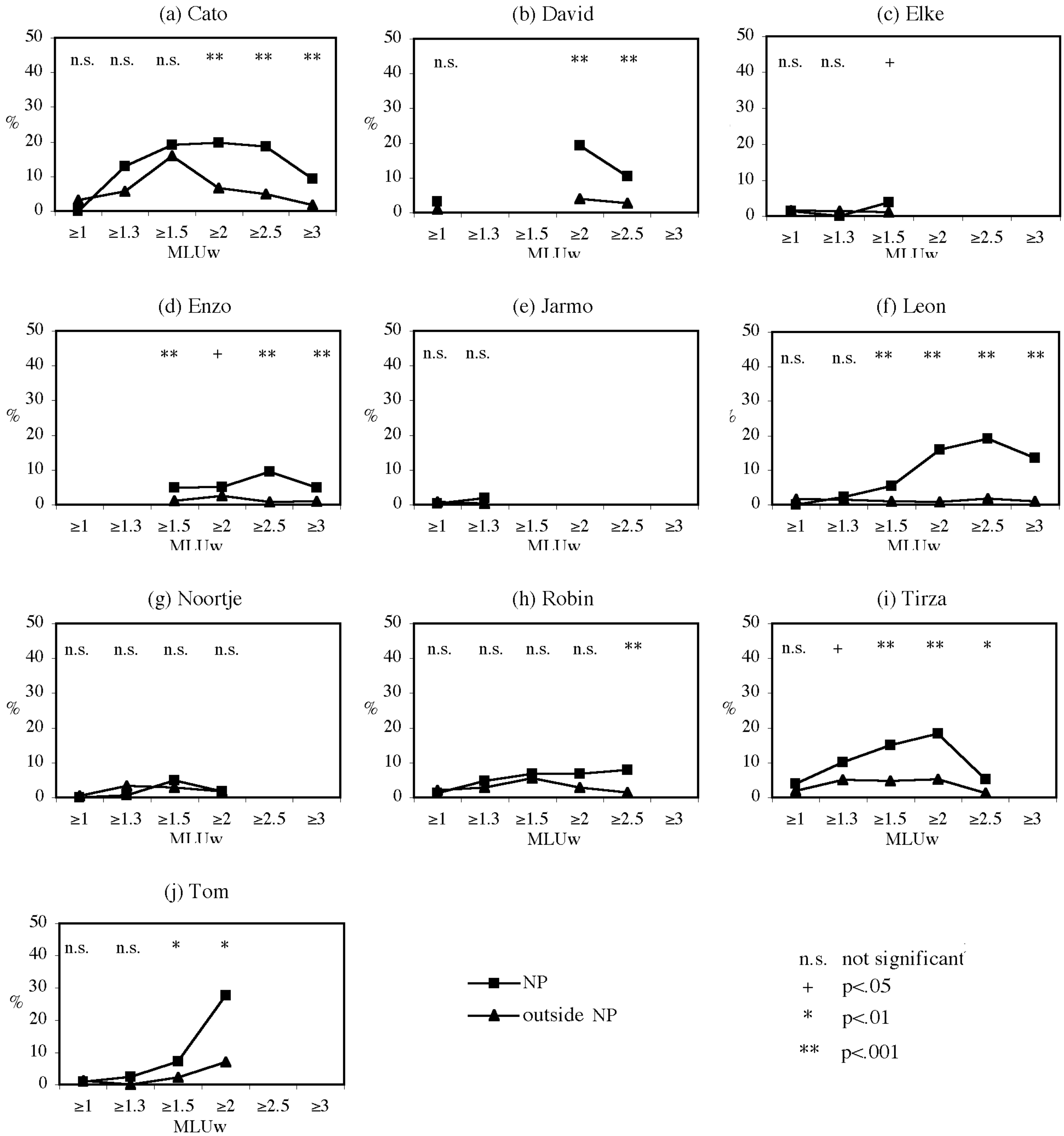


Figure 2. The frequency of schwa fillers in relation to their occurrence after anchorwords and at the beginnings of NPs

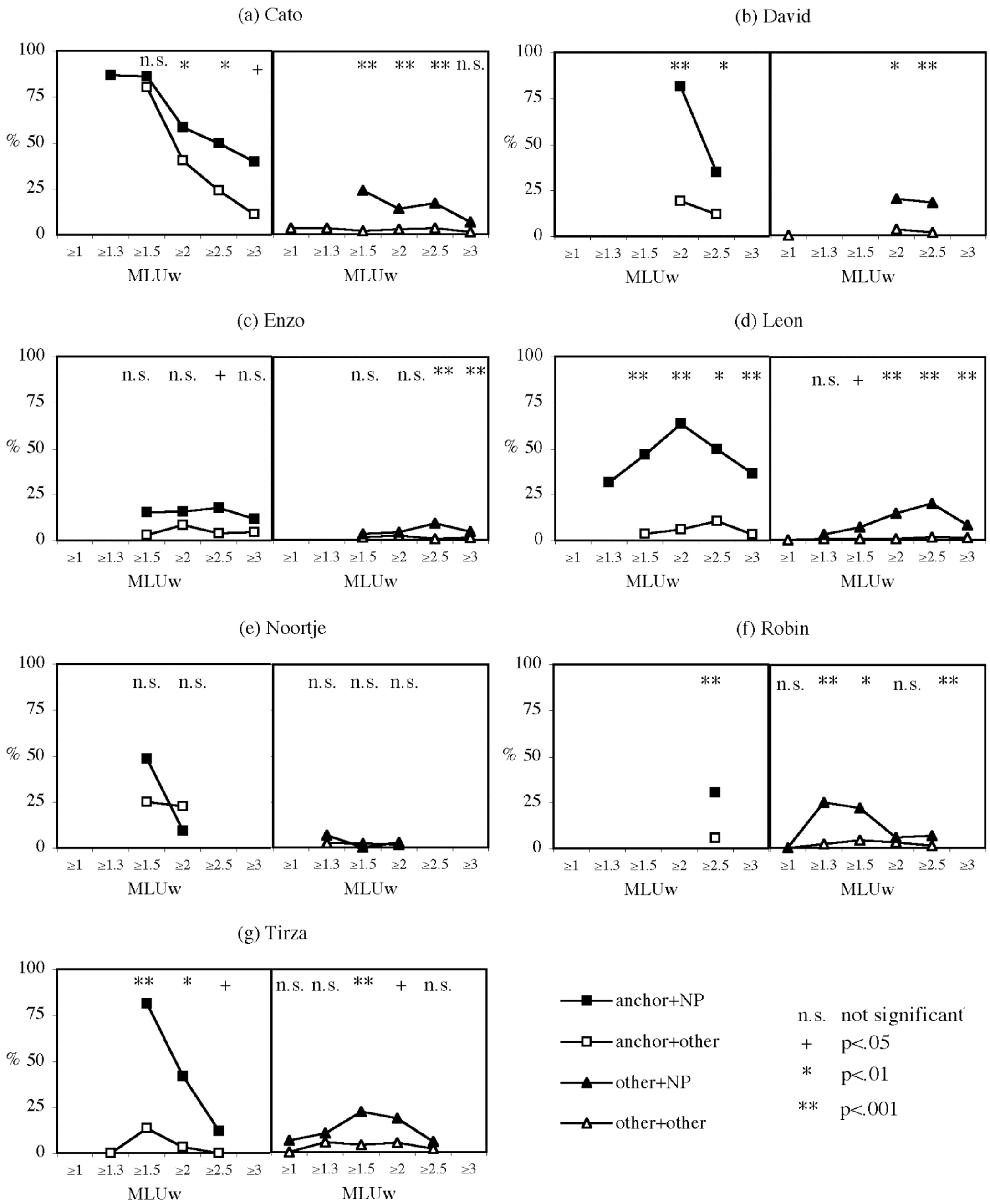


Figure 3. The determiner position over time

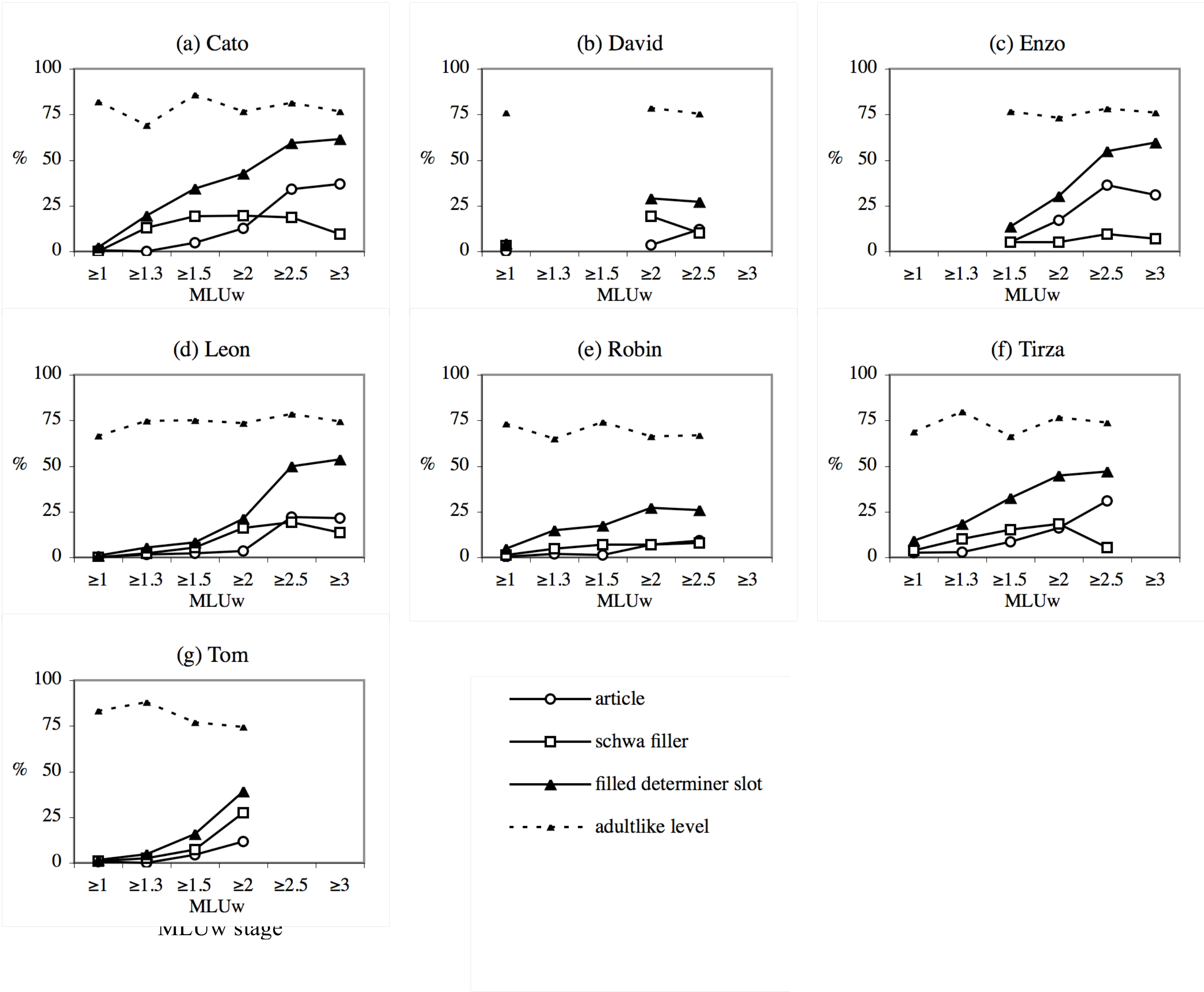
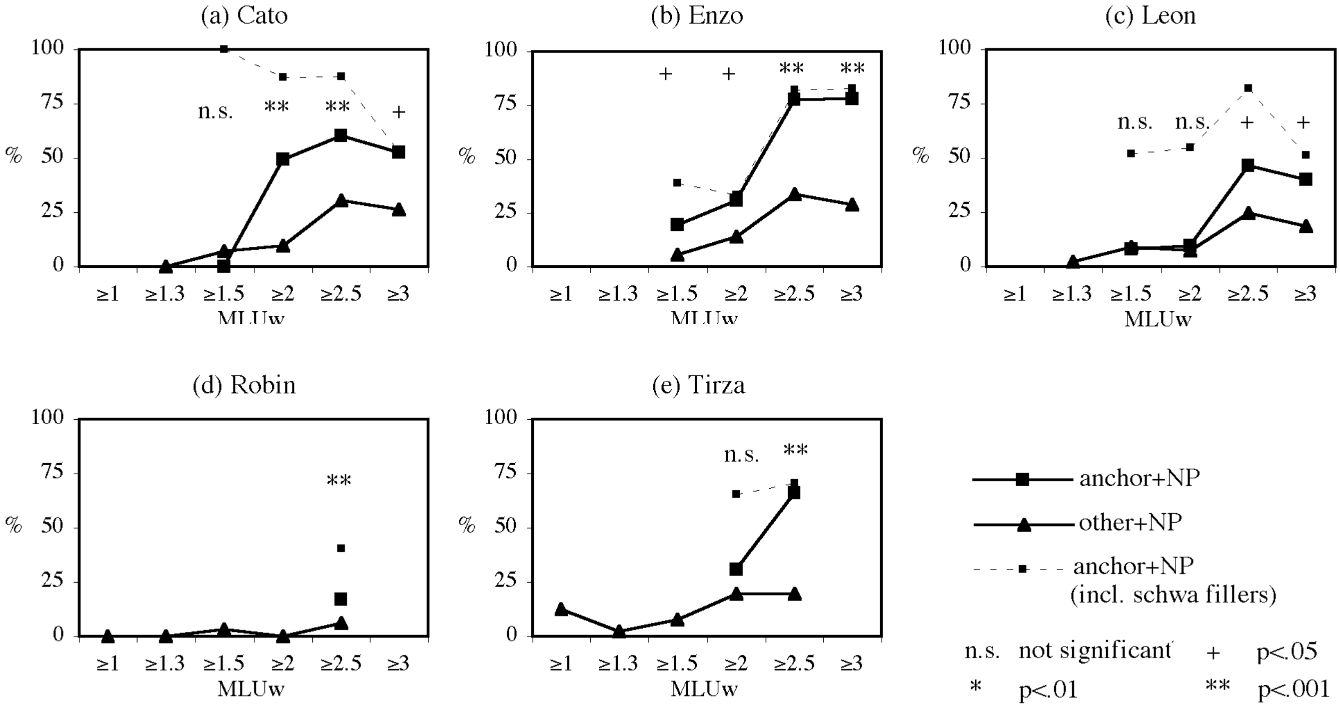


Figure 4. The frequency of articles in relation to their occurrence after anchorwords and at the the beginnings of NPs



Appendix A. Distributional analysis of the articles *een/de/het* in child directed speech

Rank	*Word*	*Frequency before articles in the input*	*Anchor word before fillers in…*	*Anchor word before articles in…*
1	is	10.1%	Cato, David, Enzo, Leon, Robin, Tirza	Enzo, Robin, Tirza
2	in	7.3%	Cato, David, Enzo, Leon, Robin, Tirza	Leon, Robin, Tirza
3	op	4.9%	Cato, David, Leon, Noortje, Robin	Enzo, Leon, Robin, Tirza
4	je	4.0%		
5	nog	3.9%	David, Leon, Tirza	Tirza
6	aan	3.3%		
7	met	3.0%		Cato, Leon, Robin
8	van	2.5%		Leon
9	ook	2.4%	Cato, David, Enzo, Leon, Noortje, Robin	Cato, Enzo, Robin
10	en	2.3%	David, Noortje	
22	gaat	1.1%	Cato	Cato, Leon
33	uit	0.6%		Robin
36	zit	0.5%		Tirza
37	niet	0.5%		Robin, Tirza
39	zijn	0.5%	Tirza	
53	heb	0.4%		Robin
71	een	0.2%	Robin, Tirza	Cato
75	dit	0.2%		Enzo

120	kijk	Noortje
1843	aardig 0.001%	

REFERENCES

Abu-Akel, A. & Bailey, A. L. (2000). Acquisition and use of ‘a’ and ‘the’ in English by young children. BUCLD 24, 45-57.

Bassano, D. & Eme, P. E. (2001). Development of noun determiner use in French children: Lexical and grammatical bases. In M. Almgren, A. Barreña, M.J. Ezeizabarrena, I. Idiazabal & B. MacWhinney (eds), *Research on Child Language Acquisition. Proceedings of the 8th Conference of the IASCL*, Somerville, MA: Cascadilla Press.

Bassano, D., Maillochon, I. & Mottet, S. (2008). Noun grammaticalization and determiner use in French children’s speech: A gradual development with prosodic and lexical influences. *Journal of Child Language* **35**, 403-438.

Bohnacker, U. (1997). Determiner phrases and the debate on functional categories in early child language. *Language Acquisition* **6**, 49-90.

Bol, G. W. & Kuiken, F. (1990). Grammatical analysis of developmental language disorders: A study of the morphosyntax of children with specific language disorders, with hearing impairment and with Down’s syndrome. *Clinical Linguistics and Phonetics* **4**, 77–86.

Bottari, P., Cipriani, P. & Chilosi, A.M. (1993/1994). Protosyntactic devices in the acquisition of Italian free morphology. *Language Acquisition* **3**, 327-369.

Dabrowska, E. (2001). Discriminating between constructivist and nativist positions: fillers as evidence of generalization. *Journal of Child Language* **28**, 243-245.

Demuth, K. & Tremblay, A. (2008). Prosodically-conditioned variability in children's production of French determiners. *Journal of Child Language* **35**, 99-127.

Dressler, W.U. & Kilani-Schoch, M. (2001). On the possible rise and inevitable fall of fillers. *Journal of Child Language* **28**, 250-253.

Feldman, A. & Menn, L. (2003). Up close and personal: A case study of the development of three English fillers. *Journal of Child Language* **30**, 735-768.

Fikkert, P. (1994). *On the acquisition of prosodic structure.* The Hague: Holland Academic Graphics.

Gerken, L. A. (1996). Prosodic structure in young children's language production. *Language* **72**, 683-712.

Gillis, S., & De Houwer, A. (1998). *The acquisition of Dutch*. Amsterdam/Philadelphia: Benjamins.

Höhle, B., Weissenborn, E., 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**, 341-353.

Haeseryn, W., Romijn, K., Geerts, G., de Rooij, J. & van den Toorn, M.C. (1997). *Algemene Nederlandse Spraakkunst.* Groningen: Wolters-Noordhoff.

Kilani-Schoch, M. & Dressler, W. U. (2001). Filler + Infinitive and pre- and protomorphology demarcation in a French acquisition corpus. *Journal of Psycholinguistic Research* **30**, 653 - 685.

Levelt, C. C. (1994). *On the acquisition of place.* Dordrecht: Holland Institute of Generative Linguistics.

Levelt, C.C. (2002) Schwa? Schma! De verwerving van schwa in de productie van kindertaal. Paper presented at the *NET symposium,* Nijmegen.

Lleo, C. (2001). Early fillers: Undoubtedly more than phonological stuffing. *Journal of Child Language* **28**, 262-265.

Lopez-Ornat, S. (2001). Fillers: How much do they generalize? *Journal of Child Language* **28**, 243-245.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk.* Hillsdale, NJ: Lawrence Erlbaum.

Pepinsky, T., Demuth, K. & Roark, B. (2001). The status of filler syllables in children's early speech. *BUCLD* **25**, 575-586.

Peters, A. M. (2001a). Filler syllables: What is their status in emerging grammar? *Journal of Child Language* **28**, 229-242.

Peters, A. M. (2001b). From prosody to grammar in English. The differentiation of catenatives, modals, and auxiliaries from a single protomorpheme. In J. Weissenborn & B. Höhle (eds), *Approaches to bootstrapping. Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition*, Amsterdam/Philadelphia: Benjamins.

Peters, A. M. & Menn, L. (1993). False starts and filler syllables: Ways to learn grammatical morphemes. *Language* **69**, 742-777.

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

Pine, J. M. & Martindale, H. (1996). Syntactic categories in the speech of young children: The case of the determiner. *Journal of Child Language* **23**, 369-395.

Rozendaal, M., & Baker, A. E. (in press). A cross-linguistic investigation of the acquisition of the pragmatics of indefinite and definite reference in two-year-olds. *Journal of Child Language*.

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

Shi, R., Werker, J. F., & Cutler, A. (2006). Recognition and representation of function words in English-learning infants. Infancy, 10, 187-198.

Taelman, H., Durieux, G. & Gillis, S. (2008) Fillers as signs of distributional learning. *Journal of Child Language* **35**, 1-31.

Theakston, A., Lieven, E., Pine, J., & Rowland, C. (2005). The acquisition of auxiliary syntax: BE and HAVE. *Cognitive Linguistics* 16, 281-311.

Tomasello, M. (2003). *Constructing a Language. A Usage-based Theory of Language Acquisition.* Cambridge, MA London, England: Harvard University Press.

Valian, V. (1986). Syntactic categories in the speech of young children. *Developmental Psychology* **22**, 562-579.

Van der Velde, M. (2003). *Déterminants et Pronoms en Néerlandais et en Français : Syntaxe et Acquisition.* Unpublished doctoral dissertation, Université de Paris-5 & Université Paris-8.

Van Kampen, N. J. (1994). The learnability of the left branch condition. In R. Bok-Bennema & C. Cremers (eds), *Linguistics in the Netherlands 1994*, Amsterdam: John Benjamins.

Veneziano, E. & Sinclair, H. (2000). The changing status of 'filler syllables' on the way to grammatical morphemes. *Journal of Child Language* **27**, 461-500.

Vollmann, R. (1997). Über die Kategorie "Fillers" im Erstspracherwerb. *Grazer Linguistische Studien* **48**, 103-132.

Wijnen, F. (1988). Spontaneous word fragmentations in children: Evidence for the syllable as a unit in speech production. *Journal of Phonetics* **16**, 187-202.

Wijnen, F. & G. Bol (1993). The escape from the optional infinitive stage. In A. de Boer, J. de Jong & R. Landeweerd (eds), *Language and Cognition 3*, Groningen: University of Groningen.

ENDNOTES

[i] *Nog* is a modal particle which can take various different meanings that can be paraphrased in English as 'more', or 'another one', or 'still' and the like. In early language development, children use it predominantly with the meaning 'another (one)' (Gillis & De Houwer, 1998).

[ii] This is hard to do since the English singular indefinite article *a* shares the form of the common schwa filler.

[iii] In the Northern Dutch section of the Spoken Dutch Corpus (http://www.tst.inl.nl; K. Luyckx, personal communication), 11% of all een tokens are realized as a schwa, and 4% are realized as a single nasal. Schwa also occurs as sloppy pronunciation variant of de and het, but less frequently (4% of de realizations, 1% of het realizations).

[iv] This percentage was obtained on the basis of an analysis of child directed speech in the Groningen database (available through CHILDES, MacWhinney, 2000). We limited the analysis to nouns that occurred in the children's vocabulary. Only realizations of *een* ('a'), *de* ('the') and *het* ('the') were counted as articles.

[v] The most frequent words after articles in the input do not reach high frequencies (yet) in the children's speech. We still believe that they may attract more schwa fillers (or articles), but controlled experimental circumstances are needed to prove this.