J. Child Lang. **36** (2009), 323–353. © 2008 Cambridge University Press
doi:10.1017/S0305000908008982 Printed in the United Kingdom

Fillers as signs of distributional learning*

HELENA TAELMAN, GERT DURIEUX
AND STEVEN GILLIS
CNTS, University of Antwerp

(*Received* 5 *October* 2006. *Revised* 23 *October* 2007. *First published online* 15 *September* 2008)

ABSTRACT

A longitudinal analysis is presented of the fillers of a Dutch-speaking child between 1;10 and 2;7. Our analysis corroborates familiar regularities reported in the literature: most fillers resemble articles in shape and distribution, and are affected by rhythmic and positional constraints. A novel finding is the impact of the lexical environment: particular function words act as 'anchor' words that attract occurrences of schwa fillers after them. The child inserts significantly more schwa fillers in these contexts. The anchor words are among the most frequent words preceding articles in the input, indicating a sharp sensitivity to such distributional regularities. Nasal fillers too are affected by distributional learning, but at the phonological level: the child first uses nasals before [h]-initial nouns, and then generalizes this usage to all [h]-initial words. These observations are related to the growing body of evidence for the impact of distributional learning on early language production.

INTRODUCTION

Young children's early utterances often contain 'fillers': vocalizations that do not correspond to conventional words. These fillers occur in various shapes: as a syllabic nasal or as a vowel, generally a schwa, sometimes also as a canonical syllable (CV), or even a string of syllables (Peters, 2001*a*). Typically, some children produce a lot of fillers, whereas they are nearly absent in the speech of other children (a.o. Peters & Menn, 1993). Moreover, they may be more frequent in some languages than in others (Peters, 1997).

[*] The research reported in this paper was enabled by a postdoctoral research grant of the National Science Foundation – FWO awarded to the first author, and a grant from the Research Council of the University of Antwerp. Address for correspondence: Helena Taelman, University of Antwerp, Dept. of Linguistics – CNTS, Prinsstraat 13, B-2000 Antwerpen, Belgium. tel: 0032-3 820.27.89; e-mail: Helena.Taelman@telenet.be

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

The interpretation of these fillers is often not straightforward: for instance, it is hard to determine which (if any) Dutch words are targeted by the fillers [n] and [ə] in utterance (1) (taken from the Maarten corpus in CHILDES, MacWhinney, 2000).

(1)	mama #	[n]	Chris	[ə]	haken	(Maarten, 1;11.4)
	mommy	F(iller)	Chris	F(iller)	crochet-INF	

Yet several in-depth studies have demonstrated that fillers are by no means without function or meaning. Instead, they prove to be highly informative for understanding children's conceptualization of the ambient language. For instance, Veneziano & Sinclair (2000) found that a French girl initially used fillers in front of monosyllabic words in order to accommodate her phrases into an iambic rhythm (weak Strong), which reflects the predominantly iambic rhythm of adult French. After a few months, the child's fillers underwent a functional reorganization. They became more frequent in prenominal position, i.e. the canonical position of determiners in French, and their phonological shapes moved towards the distribution of vowels in French determiners. Apparently they functioned as 'proto-articles' in the child's speech, i.e. elements that reveal a partial understanding of the distribution and shape of articles in the input.

The morphosyntactic and phonological determinants mentioned by Veneziano & Sinclair (2000) recur in many other studies of fillers. Fillers act as proto-articles in children acquiring French, German, English and Spanish (Bassano & Eme, 2001; Kilani-Schoch & Dressler, 2001; Lleo, 2001*a*; Lopez-Ornat, 2001; Pepinsky, Demuth & Roark, 2001; Peters & Menn, 1993). But they can also target other grammatical categories, such as modal verbs (Kilani-Schoch & Dressler, 2001; Peters, 2001*b*; Veneziano & Sinclair, 2000), question words (Dabrowska, 2001) or comparative adverbs and relative clause markers (Feldman & Menn, 2003).

Rhythm is another common factor. Children often insert fillers in order to change the original utterance rhythm into their preferred rhythmic pattern (Kilani-Schoch & Dressler, 2001; Lleo, 2001*a*; Pepinsky *et al.*, 2001; Veneziano & Sinclair, 2000). The exact content of this rhythmic preference and its onset in time appear to differ from language to language, and even from child to child (Lleo, 2001*b*; Peters, 1997). There is an ongoing debate on the question whether children start with a purely rhythmic stage of filler insertions, as in the case study of Veneziano & Sinclair (2000; also Kilani-Schoch & Dressler, 2001; Peters, 2001*a*), or whether fillers are proto-grammatical from the start, while rhythm only acts as a soft constraint on the likelihood of their production (Lleo, 2001*a*). Closely related to these rhythmic constraints are a number of positional constraints. Some fillers are typically found in utterance-initial

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

position, whereas others occur mainly in-between words (Feldman & Menn, 2003).

Apart from these morphosyntactic and phonological determinants, pragmatic factors may play a role: Feldman & Menn (2003) reported on an English boy who used multisyllabic fillers to introduce presentative and existential utterances. But we will not pursue this direction in the remainder of this paper.

In this study, we will provide a quantitative analysis of the filler productions of a Dutch girl, Cato. Since no systematic study of fillers in Dutch has been conducted yet, our analysis will break new ground in this respect. We will focus on morphosyntactic and phonological aspects of this child's filler insertions. But during the analysis it will become clear that another dimension plays a crucial role, viz. the lexical environment of fillers. We will show how fillers are tied to particular words, and we will relate these co-occurrences to distributional features of the ambient language. Finally, we will study the interaction of this lexical dimension with morphosyntactic factors and rhythmic and positional effects.

Wijnen, Krikhaar & den Os (1994) is the only study that touches upon the subject of fillers in Dutch, albeit in an indirect way. This study investigated the impact of rhythmic and positional constraints on the omission of determiners in Dutch children's utterances (cf. Gerken, 1996, for English). The prediction was that children prefer rhythmic sequences of regularly alternating stressed and weak syllables as in (2), and omit determiners if they violate this preferred rhythm.

(2) S w S w (constructed example; S = stressed, w = weak)
zoek de ganzen
find the geese

A second hypothesis was that children avoid utterances starting with a weak syllable, and hence avoid the production of an unstressed determiner in utterance-initial position. These predictions were partly confirmed in an analysis of two Dutch children's omissions of determiners. In accord with the first hypothesis, one child avoided determiners before nouns starting with a weak syllable. The other child did not. Both children produced fewer determiners in utterance-initial position than in other positions.

Wijnen *et al.* (1994) noted a high incidence of fillers in their data. As these fillers often appeared in determiner position, they decided to consider them as realizations of determiners. While this practice was certainly defendable given the aims of their study, we will disentangle a number of issues by differentiating fillers from real, full-fledged determiner productions. More specifically, we will not assume a priori that fillers are related to articles in the ambient language. Neither will we assume that

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

fillers and articles only differ in phonological form and are further functionally equivalent. Instead, we will carefully define what we consider to be fillers, verify empirically whether they are connected to articles, and re-examine the evidence on positional and rhythmic constraints for the whole group of fillers, not only for those which appear in determiner position.

METHOD

We analyzed longitudinal data of a Dutch girl, Cato. These data are part of the CLPF database, a corpus containing material on twelve children (Fikkert, 1994; Levelt, 1994; available through CHILDES, MacWhinney, 2000). Cato was recorded during seventeen naturalistic observation sessions spanning a period of ten months from age 1;10 until age 2;7. All child utterances were transcribed orthographically and phonetically, and suprasegmental features such as word stress were annotated. Since this corpus was collected with the aim of phonological analysis, the phonetic transcriptions were made carefully, and consensus with a second transcriber was aimed at (Fikkert, 1994; Levelt, 1994). Adult utterances were not included in the transcripts. The resulting corpus for Cato consists of 2243 child utterances across sessions. At the start of observations, the child's MLU in words (MLUw) was 1·1 ($SD = 0{\cdot}4$), and at the end her MLUw had increased to 3·9 ($SD = 2{\cdot}3$). Over time, Cato's cumulative vocabulary grew from 81 word types to 723 word types. For the purpose of quantitative analysis, we divided the corpus into three time slots of three/four months each: 1;10–2;0, 2;1–2–3, 2;4–2;7. This division ensured a sufficient amount of data for statistical comparison in each time slot.

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		(1;10)
	/'nɔχ	ən	'pus/		
	['nɔkə	əm	'pys]		
	more	a	cat		
(3b)	nog	F	een	poes	(1;10)
	/'nɔχ	-	ən	'pus/	
	['nɔk	ə	əm	'pys]	
	more	F	a	cat	

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

(4a)	op	tafel	(1;11)
	/ˈɔp	ˈtafəl/	
	[n	ˈtafɔ]	
	on	table	
(4b)	F	tafel	(1;11)
	/-	ˈtafəl/	
	[n	ˈtafɔ]	
	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 [ˈnɔkə] in (3a) was split into [nɔk] and [ə], and transcribed as a sequence of *nog* (/ˈnɔX/, 'more') 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 declination suffix /ə/. If the child attached a schwa to an adjective, we did not interpret this vowel as a filler, but as an attempt at the adjectival declination 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 easy, 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.[1] If we 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 is 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

[1] In the Northern Dutch section of the Spoken Dutch Corpus (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).

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

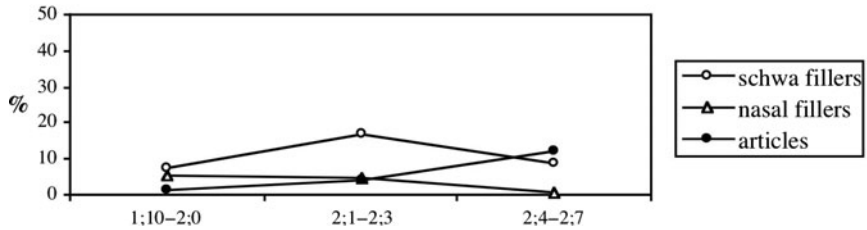


Fig. 1. Percentage of schwa fillers, articles and nasal fillers relative to the total number of words (excl. fillers and articles).

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.

Note that these criteria exclusively refer to phonological form. We decided to disregard syntactic context entirely during the annotation process, in order to be able to investigate possible syntactic effects in our analysis. Without this precaution, we would risk conflating the very issues we seek to disentangle. For the same reason, we also decided to identify realizations of the Dutch articles *een, de* and *het* independently of syntactic position: even if they did not occupy a suitable (i.e. NP-initial) grammatical position, they were interpreted as attempts at *een, de* or *het*, as long as they matched the criterion of phonological resemblance to adult articles.

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 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 to lack an obligatory determiner.

RESULTS

Cato's data contain 623 fillers and 338 articles. Most fillers consist of a schwa or a schwa-like vowel ($N=491$), but a limited group contains a nasal sound ([n] or [m], $N=132$). These nasals are syllabic in nature: 73% of them are only surrounded by consonants. Figure 1 displays the frequency of schwa fillers, nasal fillers as well as articles relative to the number of other words (all words except fillers and articles) in three time slots of three/four

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

TABLE 1. *The frequency of nasal fillers at the beginnings of NPs and outside NPs*

Age	% at beginnings of NPs	% outside NPs	Chi-squared test
1;10–2;0	5·0 ($N=695$)	6·7 ($N=195$)	X^2 (1, $N=890$) = 0·79, $p=0·37$
2;1–2;3	6·1 ($N=625$)	3·3 ($N=641$)	X^2 (1, $N=1266$) = 5·60, $p=0·02$
2;4–2;7	2·0 ($N=765$)	0·4 ($N=1373$)	X^2 (1, $N=2138$) = 11·73, $p<0·01$

months each. All frequencies are below 10% in the first time slot (age 1;10–2;0). Nasal fillers become less frequent over time, article realizations gradually become more frequent, and the frequency of schwa fillers passes through a rise and a decline.

In what follows, we will first analyze the child's usage of nasal fillers, and then the schwa fillers. For both types of fillers, we will test Wijnen *et al.*'s idea that Dutch fillers function as articles, and investigate the impact of rhythmic and positional constraints. In the section on schwa fillers, due attention will also be paid to the impact of the lexical environment. In a last section, we will compare the characteristics of fillers with those of the child's article realizations.

NASAL FILLERS

In this section we will be concerned with nasal fillers ($N=132$). First, we will investigate whether these fillers can be linked to attempts at articles; in the next subsection, we will examine the impact of positional or rhythmic constraints.

Nasal fillers and articles

If Cato's nasal fillers play the role of adult articles, we expect them to occur in the canonical position of articles, i.e. at the beginning of an NP. We tested this hypothesis by computing the percentage of nasals at the beginnings of NPs relative to the total number of NPs, and comparing this figure with the percentage of nasal fillers in other positions, relative to the total number of words outside NPs. This last part is crucial, since it takes into account the possibility that the child inserts fillers randomly, irrespective of the next word's function. All nasal fillers were included in the analysis, except those that occur next to another filler or article ($N=4$).

The frequencies of nasal fillers in both conditions are provided in Table 1, as well as a statistical test of the difference between NP-initial and NP-external positions. It appears that they are distributed randomly during the

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

TABLE 2. *The frequency of nasal fillers in utterance-initial versus utterance-internal position*

Age	% utterance initially	% utterance internally	Chi-squared test
1;10–2;0	4·9 ($N=711$)	4·9 ($N=266$)	X^2 (1, $N=977$) = 0·00, $p=0·98$
2;1–2;3	4·3 ($N=694$)	4·0 ($N=725$)	X^2 (1, $N=1419$) = 0·09, $p=0·76$
2;4–2;7	0·2 ($N=838$)	1·3 ($N=1459$)	X^2 (1, $N=2297$) = 6·65, $p=0·01$

first time slot, but that from the second time slot on, they are slightly more frequent at the beginnings of NPs. This means that Wijnen *et al.*'s impression does not apply to the early sessions, but receives positive evidence for the later ages. We will discuss later that this effect stems from a limited subset of nasal fillers.

Impact of position and rhythm on nasal fillers

Wijnen *et al.* (1994) claimed that children avoid monosyllabic and unstressed words in utterance-initial position. Additionally, they assumed that children avoid unstressed syllables next to another unstressed syllable. Since nasal fillers are monosyllabic and usually unstressed, they should rarely occur in these positions.

To investigate the first assumption, we contrasted the frequency of nasal fillers in utterance-initial position and in utterance-internal position. The former was defined as the proportion of utterance-initial nasal fillers on the total number of utterances. The latter was defined as the proportion of utterance-internal nasal fillers on the total number of utterance-internal transitions between words (where a nasal could possibly emerge). Table 2 displays these proportions in the form of percentages. The observed frequency is not higher in utterance-internal positions, except in the third time slot. We may conclude that the predicted advantage for utterance-internal position is minimally present in the data.

Does Cato avoid sequences of unstressed syllables? In other words, what is the impact of rhythm on her usage of nasal fillers? In principle, nasal fillers can occur in four utterance-internal rhythmic environments: (1) between two stressed syllables as in (5a); (2) between a weak and a stressed syllable as in (5b); (3) between a stressed and a weak syllable, as in (5c); or (4) between two weak syllables as in (5d). Wijnen *et al.*'s hypothesis predicts avoidance of nasal fillers if this would result in two or more consecutive weak syllables, i.e. scenarios (2), (3) and (4). Formulated in a positive way, Cato is expected to prefer nasal fillers between stressed syllables only, i.e. scenario (1). From this point of view, the nasal filler in (5a) is optimal, whereas the nasal fillers in (5b–d) are non-optimal, and thus avoided.

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

TABLE 3. *The frequency of nasal fillers in optimal and non-optimal rhythmic contexts*

Age	% in optimal contexts	% in non-optimal contexts	Chi-squared test
1;10–2;0	7·4 (N=148)	1·7 (N=117)	X^2 (1, N=265)=4·59, p=0·03
2;1–2;3	3·9 (N=484)	4·2 (N=237)	X^2 (1, N=721)=0·04, p=0·85
2;4–2;7	0·3 (N=985)	3·4 (N=465)	X^2 (1, N=1450)=24·03, p<0·01

(5a) S S *w* S
mam is [m:] paard (2;0)
mom is F horse

(5b) Sw *w* S
deze [ʔn] vos (2;1)
this+one F fox

(5c) S *w* wS
is [n] giraf (fictitious example)
is F giraffe

(5d) w Sw *w* wSw
n borstel [n] Catootje (2;2)
F brush F Cato-DIM

In order to test this hypothesis, we examined the stress patterns of the child's utterances. First, we identified all transitions between words in an utterance, ignoring fillers and articles. In (5a) for instance, there are two transitions: one between *mam* and *is* and another between *is* and *paard*. For each transition, we analyzed the first word's last syllable and the second word's first syllable. We computed the number of times that both syllables were stressed and the number of times that at least one of them was unstressed. Then, we counted the number of nasal fillers in optimal rhythmic conditions, i.e. when surrounded by stressed syllables, and the number of nasal fillers in non-optimal rhythmic conditions, i.e. when occurring next to one or two unstressed syllables. Finally, we computed: (1) the proportion of nasal fillers in optimal rhythmic conditions relative to the total number of times a transition consists of stressed syllables; and compared this percentage to (2) the proportion of nasal fillers in non-optimal rhythmic conditions, relative to the number of times a transition contains at least one unstressed syllable.

Table 3 shows that the ratio of optimal nasal fillers surpasses the ratio of non-optimal nasal fillers in the first time slot, that they equal each other in the second time slot, and that the non-optimal nasal fillers are more frequent in the last time slot. This means that our data do not display a clear and consistent impact of utterance rhythm on nasal fillers.

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

Phonological environments

The production of nasal fillers seems related to the phonological environment in which they occur. Half of the nasal fillers occur before words starting with a stop (66/132). More specifically, labial nasals tend to occur before labial stops, and coronal nasals before coronal stops, as in examples (6). Dorsal nasal fillers do not occur in the data.

(6a)	m	beestje	(1;11)
	/m	'bestjə/	
	[m:	'besa]	
	F	animal-DIM	
(6b)	n	jas	(1;11)
	/n	'jɑs/	
	[ʔn	'tas]	
	F	coat	

In the first time slot, 8·3% of all words starting with a labial or coronal stop ($N=338$) are preceded by such a homorganic nasal, while the frequency of nasals in front of other consonants (except [h] or glottal stop, cf. infra) is only 3·0% ($N=201$, χ^2 (1, $N=539$)$=5·99$, $p=0·01$). The frequency of homorganic nasal fillers before labial and dental stops decreases during the next two time slots to 5·0% ($N=341$) and 1·1% ($N=453$). We hypothesize that the homorganic nasals observed in the first time slot may originate from inadequate control of the velum during occlusion before the release of the stop consonant. Their subsequent decrease during the second and third time slots may then signal improving control of the velum. Similar patterns have been noted in English-speaking and German-speaking children (A. Peters, B. Lintfert, personal communication).

In the second time slot, nasal fillers frequently appear in a new context: before words starting with [h], as in (7a–c). There are 23 such instances between ages 2;1–2;3, which represent 8·9% of all [h]-initial words ($N=258$). Words starting with another consonant are preceded in only 3·8% of the cases ($N=770$) by a nasal filler during the second time slot (χ^2 (1, $N=1028$)$=10·67$, $p<0·01$). Nasal fillers also occur before [h]-initial words during the first and the third time slots, but with a lower frequency: 4·5% (10 out of 221) during the first time slot, and 2·8% during the third time slot (10 out of 361).

(7a)	n	hamer	(2;2)
	/n	'hamər/	
	[ʔn	'ɦamu]	
	F	hammer	

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

TABLE 4. *The frequency of nasal fillers before words beginning with* [h]

	Target h		Target not h		
Age	% at beginnings of NPs	% outside NPs	% at beginnings of NPs	% outside NPs	Chi-squared test[1]
1;10–2;0	19·6 (N=46)	0·0 (N=11)	0·8 (N=128)	0·0 (N=20)	Yates' X^2 (1, N=205) =23·64, p<0·01
2;1–2;3	18·9 (N=37)	15·6 (N=32)	7·6 (N=118)	3·0 (N=67)	Yates' X^2 (1, N=254) =3·81, p=0·05
2;4–2;7	29·2 (N=24)	1·4 (N=69)	1·3 (N=78)	0·6 (N=165)	Yates' X^2 (1, N=336) =15·48, p<0·01

[1] As the chi-squared tests of the frequencies in the four contexts were unreliable (due to too many expected frequencies below 5), we contrasted the frequency in the first context with the frequency in the three other contexts.

(7b) n varken (2;2)
/n 'vɑrkə/
[n 'hauku]
F pig

(7c) Roos n helpen (2;2)
/'ros n 'hɛlpə/
['hos ʔn 'ɦaupə]
Roos F help-INF

We believe that the child has derived these nasal fillers from instantiations of the article *een* in sequences of *een*+noun beginning with [h] in the input. This hypothesis is based on the fact that most nasals before [h] occur in the canonical article position at the beginning of an NP (34/43 across ages), as in (7a–b). In general, the next word is not only produced with an [h], but also has a target form starting with an [h] (29/43), as in (7a, c). It is as yet unclear why the nasal in *een* was more prominent to the child than the schwa before [h]-initial nouns.

The burst of nasal fillers before [h]-initial words in the second time slot is caused by the expansion of this pattern outside its original context: now, 7 instances occur outside an NP, and in 11 instances a nasal filler precedes an [h]-initial rendition of a target word (which has no initial [h] in the adult model). It seems as if the child has interpreted the original pattern as a phonological regularity applicable to all [h]-initial words. Table 4 displays the percentages of nasal fillers before [h] relative to the total number of [h]-initial words in four contexts: (1) before words with [h]-initial targets, at the beginnings of NPs; (2) before words with [h]-initial targets, outside NPs; (3) before words without [h]-initial targets, at the beginnings of NPs; and (4) before words without [h]-initial targets, outside NPs. In the first time

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

TABLE 5. *The frequency of nasal fillers at the beginnings of NPs and outside NPs after exclusion of nasal fillers before words beginning with* [h]

Age	% at beginnings of NPs	% outside NPs	Chi-squared test
1;10–2;0	3·6 ($N=695$)	6·7 ($N=195$)	X^2 (1, $N=890$) = 3·51, $p=0·06$
2;1–2;3	3·5 ($N=625$)	2·2 ($N=641$)	X^2 (1, $N=1266$) = 2·04, $p=0·15$
2;4–2;7	0·9 ($N=765$)	0·4 ($N=1373$)	Yates' X^2 (1, $N=2138$) = 1·78, $p=0·18$

slot, nasal fillers are only frequent in the first context, i.e. at the beginnings of NPs and before words with [h]-initial targets. In the second time slot, nasal fillers often occur in the other contexts as well. By the third time slot, the child has restricted these nasals to the first context again.

Although this specific observation has never been described before, similar phonological interpretations of morphosyntactic phenomena have been signalled in the literature, often as idiosyncratic as in our case (Peters & Menn, 1993; Vihman, 1996: 225). For instance, the English-speaking boy Daniel interpreted the English [z]-allomorph on a phonological rather than a morphological basis (Peters & Menn, 1993). While this morpheme occurs as a plural suffix, possessive suffix, contracted auxiliary or copula in adult English, he disregarded these morphosyntactic functions and inserted a word-final [z] depending on the word's phonological form (when two-syllabic and ending in [r], or having a medial coronal and ending in [i]). Peters & Menn speculated that this phonological generalization was based on some frequent plurals and possessives in the input with these specific phonological shapes. In the case of Cato, we see a two-step process, where, for some reason, the child first translates *een*+[h]-initial noun sequences into nasal+[h]-initial nouns, and then becomes aware of the phonological pattern in these sequences through distributional learning and extends it to phonologically similar contexts.

To sum up, phonemic factors guide the insertion of nasal fillers: the majority of nasal fillers precede words beginning with a stop, another set of nasal fillers precedes words beginning with [h]. Together, these factors cover 83% of all nasal filler insertions. The second factor goes together with a bias for NP-initial positions: most nasal fillers preceding [h]-initial words occur before nouns beginning with [h]. The noun bias in this subset of nasal fillers is entirely responsible for the preference for NP-initial positions that was observed in the general analysis of Table 1. Table 5 contains a recalculation of this analysis after exclusion of the nasal fillers before [h]-initial words, and shows no difference between the two conditions. This means that Wijnen *et al.*'s intuition about the link with articles only applies to a subset of nasal fillers preceding [h]-initial words.

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

TABLE 6. *The frequency of schwa fillers at the beginnings of NPs and outside NPs*

Age	% at beginnings of NPs	% outside NPs	Chi-squared test
1;10–2;0	7·5 ($N=695$)	7·7 ($N=195$)	X^2 (1, $N=890$) = 0·01, $p=0·92$
2;1–2;3	22·2 ($N=625$)	9·8 ($N=641$)	X^2 (1, $N=1266$) = 36·35, $p<0·01$
2;4–2;7	16·5 ($N=765$)	4·1 ($N=1373$)	X^2 (1, $N=2138$) = 96·87, $p<0·01$

SCHWA FILLERS

We will now adopt the same analysis techniques for the schwa fillers ($N=491$). As before, we will examine whether these fillers can or should be considered as imperfect realizations of articles, and investigate the role of positional and rhythmic constraints. Furthermore, we will have an eye for additional determinants, which will now be situated on the level of the lexical environment instead of the phonological environment. For reasons of exposition, we will present the analysis of the lexical environment before the discussion of positional and rhythmic constraints, since much of what we have to say about position and rhythm relies on it.

Schwa fillers and articles

We will examine the possible link of schwa fillers to articles in the same manner as we did for the nasal fillers, i.e. by a quantitative analysis of filler occurrence. We computed the frequency of schwa fillers at the beginnings of NPs as the percentage of schwa fillers at the beginnings of NPs over the total number of NPs, and compared this figure with the frequency of schwa fillers outside NPs estimated as the percentage of schwa fillers outside NPs relative to the total number of words outside NPs. All schwa fillers were included in the analysis ($N=451$), except for cases where a filler occurs next to another filler or article ($N=40$). Table 6 contrasts the frequencies of schwa fillers at the beginnings of NPs and outside NPs for all three time slots. It turns out that the frequencies are about equal in the first time slot, but that from age 2;1–2;3 on, the likelihood of schwa fillers starts to differentiate: the frequency of schwas outside NPs is 12% lower than at the beginnings of NPs. However, less frequent does not mean infrequent: the frequency of schwas outside NPs is still 9·8% in the second time slot and 4·1% in the third time slot. We therefore conclude that the distribution of these schwa fillers does not entirely conform to the distribution of articles, but that it is moderately skewed towards it from the second time slot onwards.

The distribution of schwa fillers differs in more than one respect from that of Dutch articles in general. The most striking difference is their

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

occurrence outside NPs. This happens in 138 instances (out of 451), six of which are displayed in (8).

(8a)	lepel ook ə in pap	(1;11)
	spoon also F in porridge	
(8b)	niet ə op ə stenen lopen	(2;4)
	not F on F stones run-INF	
(8c)	tijger (i)s ə wakker	(2;4)
	tiger is F awake	
(8d)	gaat ə fietsen	(2;3)
	will F bike-INF	
(8e)	(i)s ə spelen gitaar	(2;1)
	is F play-INF guitar	
(8f)	kunt ə goed zien	(2;3)
	can-3SG F well see	

But even in NP-initial position, schwa fillers sometimes violate restrictions of article use in adult Dutch. For instance, a number of schwa fillers ($N=9$) immediately precede proper nouns, as in (9), a context that prohibits articles.

(9a)	is ook ə Roos huilen.	(2;2)
	is also F name cry-INF	
(9b)	kijk, net als ə Catootje.	(2;4)
	look, just like F name-DIM	

Furthermore schwa fillers occur before full-fledged articles or other determiners, as in (10), although such doublings of determiners are not allowed in adult Dutch. Over time, this happens eight times.

(10a)	(i)s ə ook ə de krokodil	(2;2)
	is F also F the crocodile	
(10b)	nog ə een poes	(1;10)
	also F a cat	
(10c)	kijk pop met ə z'n vingers	(2;4)
	look doll with F his fingers	

Summarizing our results, we have found that the distribution of schwa fillers tends towards that of Dutch articles from the second time slot onwards. Given their phonological similarity to articles, it is tempting to consider them as reduced forms of articles. Nevertheless, schwa fillers differ from Dutch articles in an important number of ways: unlike articles they occur outside NPs, appear before proper nouns and turn up before full-fledged articles or other determiners. These differences suggest that schwa fillers are not merely deformed attempts at adult-like articles, but that they

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

reflect the child's immature understanding of the distribution of articles in adult Dutch.

The impact of lexical environment

Before we turn to rhythmic and positional constraints on the occurrence of schwa fillers, we will now describe what is perhaps the most striking characteristic of Cato's schwa fillers: even to the casual observer of Cato's data, it will be clear that her schwa fillers are often tied to specific words. This is particularly true for the schwa fillers in the first time slot. Example (11) contains four utterances with schwa fillers produced at age 2;0. It immediately catches the eye that all schwa fillers in these utterances occur after the same inflected verb form: *is* 'is'.

(11a)	*is* ə kapot	(2;0)
	is F broken	
(11b)	donker *is* ə boven	(2;0)
	dark is F upstairs	
(11c)	dit *is* ə paard	(2;0)
	this is F horse	
(11d)	*is* ə letter	(2;0)
	is F character	

This is a typical characteristic of the schwa fillers in the first time slot: more than half of them follow the verb form *is* (41 out of 72). Six out of the remaining 31 schwa fillers occur after the word *ook* 'also'. Other words are produced at most three times before a schwa filler. Moreover, the incidence of a schwa filler after *is* and *ook* is higher than can be expected by chance: *is* is followed by a schwa filler in 75·9% of all tokens ($N=54$); *ook* is followed by a schwa filler in 66·7% of all tokens ($N=8$). Discarding *is* and *ook*, the incidence of schwa fillers after other words (excluding fillers, articles and false starts) is only 6·4% ($N=204$).[2]

In the second time slot, the sequence *is*+schwa filler remains frequent: more than half of the schwa fillers are found after that particular word ($N=131$ out of 224 schwa fillers). *Ook* remains the second most frequent preceding word ($N=15$), joined by *op* ('on', $N=12$), *gaat* ('goes', $N=11$), and *in* ('in', $N=7$). No other words precede a schwa filler more than five times (apart from the article *een* – cf. infra, analysis of example 15c). Figure 2 plots the frequency with which *is*, *ook*, *op*, *gaat* and *in* are followed by a schwa filler relative to the total number of tokens of each word.

[2] The data do not enable a fine-grained analysis of the incidence of schwa fillers after each word because of the low number of tokens of most words. Hence, we use the percentage of schwa fillers after all other words over the total number of occurrences of these words as an estimate of their usual incidence.

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

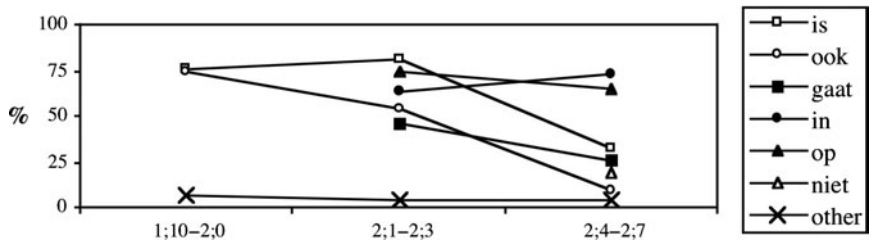


Fig. 2. The frequency of schwa fillers after *is*, *ook*, *gaat*, *in*, *op*, *niet*, and after other words.

Between age 2;1 and 2;3 all frequencies for the mentioned words are higher than 40%. This is much higher than the general frequency of the schwa filler after the remaining words ('other').

By the third time slot, 33 out of 195 tokens of schwa fillers are preceded by *is*, 29 tokens are preceded by *in*, and 27 tokens by *op*. Another anchor word is *niet* ('not', $N=6$). The word *ook* now only occurs four times in front of a schwa. Figure 2 reveals that the incidence of schwa fillers after these words is again higher than the incidence of schwa fillers after other words. For four words, *op, ook, is* and *gaat*, the frequency of schwa fillers has decreased relative to the preceding period. In the section on article realizations, we will see that attempts at articles start to occupy the place of the schwa fillers in these contexts.

The words *is, ook, op, in, gaat, niet*, will be called 'anchor words' henceforth on the basis of their attraction for schwa fillers. Table 7 reveals that these anchor words not only attract schwa fillers at the beginnings of NPs, but also outside NPs. The table contains the frequency of schwa fillers after the six anchor words on the total number of occurrences of these anchor words in contexts where anchor words precede an NP, and contexts where another syntactic constituent follows. Under both conditions, the frequency of schwa fillers is above 50% until age 2;3.

The high probability of schwa fillers after anchor words appears to be responsible for the majority of schwa fillers outside NPs: anchor words precede 73·3% of all schwa fillers outside NPs (99/135). Moreover, these anchor words often precede sequences of a schwa and a proper noun (7 out of 9 instances), and occur in front of sequences of schwa followed by a determiner (4 out of 8). As a result, the divergence between the distribution of schwa fillers in Cato's speech and the distribution of articles in adult Dutch can be partly explained by the higher probability of schwa fillers after anchor words.

This does not imply that the child completely ignores the syntactic context after anchor words: Table 7 shows that schwas are always more frequent if an NP follows. During the last two time slots, this difference is

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

TABLE 7. *The impact of the anchor words on the frequencies of schwa fillers in and outside NPs*

	After anchor words		After other words		
Age	% at beginnings of NPs	% outside NPs	% at beginnings of NPs	% outside NPs	Chi-squared test
1;10–2;0	73·6 ($N=53$)	68·8 ($N=16$)	10·0 ($N=50$)	5·4 ($N=93$)	X^2 (3, $N=212$) $=98{\cdot}79$, $p<0{\cdot}01$
2;1–2;3	78·2 ($N=156$)	59·3 ($N=91$)	12·5 ($N=88$)	2·7 ($N=329$)	X^2 (3, $N=664$) $=342{\cdot}32$, $p<0{\cdot}01$
2;4–2;7	44·9 ($N=187$)	22·8 ($N=149$)	12·8 ($N=204$)	2·8 ($N=779$)	X^2 (3, $N=1319$) $=259{\cdot}37$, $p<0{\cdot}01$

around 20% (second time slot: X^2 (1, $N=247$) $=9{\cdot}99$, $p<0{\cdot}01$; third time slot: X^2 (1, $N=336$) $=17{\cdot}78$, $p<0{\cdot}01$). At age 1;10–2;0, the difference is smaller and insignificant (Yates' X^2 (1, $N=69$) $=0{\cdot}00$, $p=0{\cdot}95$). We may conclude that from age 2;1–2;3 on, the likelihood of a schwa after an anchor word is greater when an NP follows.

The last columns of Table 7 compare the proportion of schwa fillers at the beginnings of NPs and the proportion of schwa fillers before other syntactic constituents, after any other words except the six anchor words. Now percentages are much lower than after anchor words. Also the difference between fillers at the beginnings of NPs and before other syntactic constituents is smaller, but remains significant, at least in the last two time slots (first time slot: Yates' X^2 (1, $N=143$) $=0{\cdot}48$, $p=0{\cdot}49$; second time slot: Yates' X^2 (1, $N=417$) $=12{\cdot}44$, $p<0{\cdot}01$; third time slot: X^2 (1, $N=983$) $=34{\cdot}26$, $p<0{\cdot}01$). We therefore conclude that Cato prefers to insert schwa fillers in NP-initial positions in each lexical context.

Distribution of articles in the input

Now that we have established the existence of anchor words, which attract schwa fillers, the question remains why they do so. Our hypothesis is that those sequences are derived from characteristics of the ambient language. More specifically, we will show that the highly frequent sequences of anchor words and schwa fillers correspond to highly frequent sequences of words and articles in the input.

Speech addressed to young children tends to be highly repetitive and formulaic (Cameron-Faulkner, Lieven & Tomasello, 2003). For instance, Dutch parents use sentences like (12a) and (12b) many times.

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

(12a) Dit is een muis/koe/paard/etc.
This is a mouse/cow/horse/etc.
(12b) Waar is de kikker/auto/radio/etc.
Where is the frog/car/radio/etc

A consequence of this characteristic may be that articles frequently occur after the same words e.g. after *is*, as in *dit is een X* 'this is a X', or *waar is de X* 'where is the X'. As schwa is a phonological element shared by all Dutch articles (/ən/, /də/, /(h)ət/), a child may well develop strong memory associations between such words and schwa. We believe that these sequences are not simply stored as unanalyzed holophrases as a result of incomplete segmentation. Rather, the words in such sequences are stored separately, but are tightly interlinked due to the high probabilities of co-occurrence in the ambient language. This qualification is crucial in order to account for the impact of the next word's syntactic status on the likelihood of schwa insertion. If schwa is considered as a separate word, it is easy to imagine that its occurrence is not only determined by associations with preceding words, but also by its association with succeeding NPs. If we adopt the view that the anchor words form holophrases together with schwa fillers, we would need to stipulate that the probability of the entire holophrase depends on the syntactic status of the next words. Although this is not impossible, in doing so, we would miss the generalization that schwa fillers not only occur more frequently in front of NPs when they are part of these holophrases, but also in other contexts.

Our hypothesis predicts that the sequences of anchor words plus schwa fillers in Cato's speech reflect highly frequent sequences of anchor words plus articles in adult speech. In order to test this prediction, we analyzed the distribution of words around articles in child-directed speech (abstracting away the words' syntactic functions). Since the Cato data do not contain transcriptions of the ambient language, we relied on child-directed speech from four other databases available through CHILDES: the Groningen database (Wijnen & Bol, 1993), the Van Kampen database (Van Kampen, 1994), the Wijnen database (Wijnen, 1988) and the section for normally developing children in the Bol & Kuiken database (Bol & Kuiken, 1990). The aggregated data contain 270960 adult utterances and include 79262 articles. 1843 different word forms (each inflectional variant is counted as a separate word) precede articles, but some of these occur far more frequently than others. The word *is* 'is' occurs most frequently in front of articles: *is* precedes 10% of all utterance-internal articles ($N = 71099$). The second most frequent word before articles is *in* (7·3% of all utterance-internal articles), followed by *op* (4·9% of all utterance-internal articles). These three words, *is, in* and *op*, belong to the set of anchor words in Cato's speech. The other anchor words *ook, gaat* and *niet* have a lower ranking, but are still

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

TABLE 8. *The impact of utterance position on the frequency of schwa fillers inside and outside NPs (excluded are all contexts before anchor words)*

	Beginnings of NPs		Outside NPs		
Age	% utterance initially	% utterance internally	% utterance initially	% utterance internally	Chi-squared test
1;10–2;0	1·4 (N=592)	10·0 (N=50)	0 (N=119)	5·4 (N=93)	(too many expected frequencies below 5)
2;1–2;3	1·6 (N=381)	12·5 (N=88)	0 (N=313)	2·7 (N=329)	X^2 (3, N=1111) =48·45, p<0·01
2;4–2;7	4·3 (N=374)	12·8 (N=204)	0 (N=464)	2·8 (N=779)	X^2 (3, N=1821) =69·90, p<0·01

located in the upper part of the list: *ook* is the ninth most frequent word before an article in adult speech (2·4%), *gaat* is the twenty-second most frequent word (1·1%) and *niet* is the thirty-seventh most frequent word (0·5%). Together, the six anchor words precede 26·4% of all utterance-internal (i.e. non-initial) articles in adult Dutch as spoken to children. It is highly probable that the child has picked up these strong distributional regularities. The absence of a perfect match between this input analysis and Cato's output may be explained by idiosyncrasies in the child's own input, and by her individual path of lexical acquisition.

Impact of rhythm and position on schwa fillers

Finally, what is the impact of rhythm and utterance position on the likelihood of schwa fillers? To start with the latter: are schwa fillers avoided at the start of an utterance? Table 8 displays the relative frequencies of schwas in utterance-initial positions at the beginnings of NPs and outside NPs, and compares them with the relative frequencies in utterance-internal positions (except after anchor words). At the beginnings of NPs the relative frequency is up to 10% higher utterance internally than utterance initially (second time slot: Yates' X^2 (1, N=469)=21·40, p<0·01; third time slot: X^2 (1, N=578)=14·04, p<0·01), in other contexts it is 3–6% higher (second time slot: Yates' X^2 (1, N=642)=6·82, p<0·01; third time slot: X^2 (1, N=1243)=13·34, p<0·01). Thus, corroborating Wijnen *et al.*'s hypothesis, Cato indeed avoids schwa fillers at the start of an utterance.

In utterance-internal positions, we also compared the relative frequency of schwas in optimal rhythmic conditions, i.e. between stressed syllables, with the relative frequency of schwas in non-optimal rhythmic conditions, next to one or two weak syllables, using the method described in our analysis of rhythmic constraints in nasal fillers. If Cato were avoiding

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

TABLE 9. *Percentage of schwa fillers in rhythmically optimal and rhythmically non-optimal utterance-internal positions*

Age	% in optimal contexts	% in non-optimal contexts	Chi-squared test
1;10–2;0	39·2 ($N=148$)	1·7 ($N=117$)	X^2 (1, $N=265$) = 52·41, $p<0·01$
2;1–2;3	37·2 ($N=484$)	7·6 ($N=237$)	X^2 (1, $N=721$) = 69·95, $p<0·01$
2;4–2;7	16·5 ($N=985$)	1·4 ($N=465$)	X^2 (1, $N=1450$) = 71·42, $p<0·01$

TABLE 10. *The impact of rhythm on the frequency of schwa fillers inside and outside NPs* (*excluded are all contexts before anchor words*)

	At the beginnings of NPs		Outside NPs		
Age	% in optimal contexts	% in non-optimal contexts	% in optimal contexts	% in non-optimal contexts	Chi-squared test
1;10–2;0	18·5 ($N=27$)	0 ($N=22$)	8·9 ($N=45$)	2·1 ($N=48$)	(too many expected frequencies below 5)
2;1–2;3	21·2 ($N=52$)	0 ($N=35$)	3·4 ($N=178$)	2·0 ($N=150$)	(too many expected frequencies below 5)
2;4–2;7	15·7 ($N=153$)	4·4 ($N=45$)	4·3 ($N=440$)	0·9 ($N=337$)	X^2 (3, $N=975$) = 49·94, $p<0·01$

sequences of two or more consecutive weak syllables, schwa syllables would be less frequent in the latter context than in the former.

The data in Table 9 indicate that this is indeed the case. The ratio of schwa fillers in optimal rhythmic conditions is much higher than the ratio of schwa fillers in non-optimal rhythmic conditions, with a difference of up to 37·5%. However, further analysis reveals that the high incidence of schwa fillers in optimal position is not only due to the rhythmic environment, but also brought about by the fact that many of these fillers occur after one of the anchor words. There is a conspiracy between these two factors: almost all schwa fillers after anchor words optimize the utterance rhythm (327/344 cases), and most anchor words provide an environment of two stressed syllables which is amenable to optimization by means of filler insertion (617/652 cases).

We therefore expect the impact of rhythmic constraints to be weaker outside the contexts provided by anchor words, and this prediction is borne out by the data. Table 10 displays the percentages of utterance-internal fillers in optimal and non-optimal rhythmic conditions after excluding the contexts with anchor words from the analysis. The table also distinguishes

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

between NP-initial position and other positions. The difference between optimal and non-optimal rhythmic contexts never approximates the level found in Table 9. At the beginnings of NPs, schwa fillers are up to 21·2% more frequent in optimal rhythmic contexts. They do not occur in non-optimal rhythmic contexts until the third time slot, when they are only 11·3% less frequent than in optimal rhythmic contexts (third time slot: X^2 (1, $N=198$) = 3·85, $p=0·05$). The difference between these two conditions is even smaller outside NPs, but still significant in the third time slot (X^2 (1, $N=777$) = 8·15, $p<0·01$).

We may conclude that fillers are less frequent at the start of an utterance, and in non-optimal rhythmic conditions (in the environment of at least one weak syllable). When an ideal rhythmic context co-occurs with the presence of an anchor word, this conspiracy yields a very high number of schwa fillers.

ARTICLES

Now that we have reviewed both nasal and schwa fillers, and discussed their relation to Dutch articles, we will turn to article realizations by the child ($N=338$). By the second time slot, Cato already produces a fair number of articles ($N=42$, 4·4% of the total number of word tokens) while she is still using schwa fillers ($N=202$, 16·9% of the total number of word tokens, see). Is the distribution of these forms adult-like? Is their production subject to positional and/or rhythmic constraints, and do anchor words play a role?

Comparison with articles in adult Dutch

In adult speech the articles *een* 'a', *de* 'the' and *het* 'the' almost always occur in the determiner position of NPs. In addition, adults also use *een* and *het* as personal or indefinite pronouns. In Cato's data attempts at *een, de* and *het* were identified solely on the basis of their phonological form, irrespective of syntactic considerations (see Method section). Thus, although the occurrences of *een, de* and *het* will be referred to as 'article realizations', this does not necessarily imply that they are identical to adult articles in every respect. Nevertheless, empirical analysis will reveal that their distribution approximates the adult distribution rather well.

A first indication of the similarity of Cato's article realizations to their adult counterparts is given by the degree to which her articles appear in determiner position. Table 11 displays the percentage of articles at the beginnings of NPs over the total number of NPs and the percentage of articles outside NPs over the total number of words outside of NPs. Outside NPs, their frequency is consistently low throughout the whole period,

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

TABLE 11. *Frequency of articles at the beginnings of NPs and outside NPs*

	% at beginning of NPs	% outside NPs	Chi-squared test
1;10–2;0	1·4 ($N=695$)	0 ($N=195$)	Yates' X^2 (1, $N=890$) = 1·69, $p=0·19$
2;1–2;3	5·8 ($N=625$)	0·9 ($N=641$)	X^2 (1, $N=1266$) = 22·96, $p<0·01$
2;4–2;7	29·2 ($N=765$)	1·2 ($N=1373$)	X^2 (1, $N=2138$) = 384·08, $p<0·01$

whereas there is a marked increase of NP-initial articles from 1·4% in the first time slot to 29·2% in the third time slot.

This last stage does not represent adult-like proficiency yet. Cato tends to omit determiners even in contexts where they are required in adult Dutch: in these obligatory contexts, 33·9% of all NPs ($N=533$) are produced without an article, another determiner or a filler. Nevertheless, her production of determinerless NPs has decreased drastically relative to the first time slot, when the omission rate was 82·4% in obligatory contexts ($N=552$). It seems then that over time, Cato starts to implement the canonical structure of an NP in adult Dutch, but that ample room for improvement remains.

On the other hand, Cato appears to know the circumstances that prohibit articles rather well already. Whereas schwa fillers were still relatively frequent outside NPs (>4%), articles hardly occur in these contexts: Table 11 indicates percentages below 1·5%. Note that a few of those cases are grammatical, as they concern the use of *het* as a pronoun ($N=5$). Ungrammatical insertions as in (13) only occur eighteen times, eight of which can be analyzed as false starts, which are retraced in the remainder of the utterance, as in (13b). Moreover, unlike the schwa fillers discussed earlier, but conforming to adult Dutch, articles do not occur in front of proper nouns (except once).

(13a) gaat een eten (2;4)
goes a eat-INF

(13b) heeft een niet een kroontje om (2;5)
has a not a crown-DIM on

There are two gaps in Cato's mastery of the adult distribution; both are instances of the use of the singular indefinite article *een*. This is the most frequent article in Cato's speech (310/338), and generally appears before singular nouns, conforming to adult Dutch. But unlike adult Dutch, twenty-one occurrences of *een* can be found before a plural noun, as in (14).

(14a) ik ook een blokken (2;3)
I also a block-PL

(14b) een koeien (2;5)
a cow-PL

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

(14c) gaat een liedjes zingen (2;5)
goes a song-DIM-PL sing-INF

Another puzzling use of *een* is that it sometimes precedes another determiner, as in examples (15a–b) or a schwa filler, as in (15c). The first type of construction occurs fourteen times, the second type thirty-two times.

(15a) een drie vliegtuigen (2;7)
a three plane-PL
(15b) was een zo'n vlinder (2;7)
was a such-a butterfly
(15c) een ə bus (2;1)
a F bus

Summing up, Cato's use of articles closely resembles the distribution of adult Dutch articles: in contrast to her schwa fillers, Cato's articles rarely occur outside NPs or in front of proper nouns. So, it seems that the discovery of the full phonological form accompanies a better understanding of its distribution in adult speech. Nevertheless, a few non-adult-like uses of *een* (in front of a plural noun or of another determiner) indicate that the child's grammatical knowledge is still not fully adult-like. Moreover, even at the end of the observation period, the child has not yet fully mastered the rules of obligatory determiner use.

Impact of position, rhythm and anchor words on article realizations

Let us now investigate to what extent the factors that determined the likelihood of schwa fillers also influence the likelihood of article realizations. First we will examine the possible role of rhythm and position, and then we will turn to the anchor words.

Rhythm and position will be considered together: following the methodology outlined in our analyses of fillers, we computed the percentage of articles in NPs: (1) at the start of an utterance; (2) in utterance-internal position, when surrounded by stressed syllables (the rhythmically optimal context); and (3) in utterance-internal positions next to a weak syllable (the rhythmically non-optimal context). To avoid possible interference with the impact of anchor words, NPs that occur immediately after an anchor word were excluded from the analysis.

In our analysis of schwa fillers, we found a low incidence of fillers in utterance-initial position (0–4·3%), an equally low incidence in non-optimal utterance-internal positions (0–4·4%) and the highest frequency rates in optimal utterance-internal positions (15·7–21·1%). Table 12 shows a slightly different constellation. Unlike the schwa fillers, articles in utterance-initial

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

TABLE 12. *The impact of position and rhythm on the frequency of articles (at the beginnings of NPs)*

Age	% utterance initially	% in internal and optimal contexts	% in internal and non-optimal contexts	Chi-squared test
1;10–2;0	0·7 ($N=592$)	18·5 ($N=27$)	0 ($N=22$)	(too many expected frequencies below 5)
2;1–2;3	4·2 ($N=381$)	19·23 ($N=52$)	2·9 ($N=35$)	(too many expected frequencies below 5)
2;4–2;7	25·9 ($N=374$)	37·3 ($N=153$)	13·3 ($N=45$)	X^2 (2, $N=572$) = 12·10, $p<0·01$

TABLE 13. *The frequency of articles after anchor words or after other words (restricted to utterance-internal, optimal contexts)*

	% after anchor words	% after other words	Chi-squared test
1;10–2;0	1·9 ($N=52$)	18·5 ($N=27$)	Yates' X^2 (1, $N=79$) = 4·81, $p=0·03$
2;1–2;3	6·2 ($N=146$)	19·2 ($N=52$)	Yates' X^2 (1, $N=198$) = 6·12, $p<0·01$
2;4–2;7	34·1 ($N=185$)	37·3 ($N=153$)	X^2 (1, $N=338$) = 0·37, $p=0·54$

position have become quite frequent (25·9%) by the third time slot, and considerably more so than utterance-internal non-optimal articles (13·3%). But both contexts attract fewer articles than utterance-internal optimal positions (37·3% in the third time slot). We can infer that the impact of rhythm is approximately the same, while the constraint against utterance-initial positions seems to be gradually relieved.

Our next analysis concerns the impact of the anchor words *is, ook, gaat, in, op, niet*. It will be limited to rhythmically optimal positions, since other occurrences are extremely rare. Table 13 compares the percentage of NP-initial articles following one of the six anchor words with the percentage of NP-initial articles in other contexts.

Based on what was found for schwa fillers, the incidence of articles after anchor words is expected to be much higher than that in other contexts. This, however, is not what Table 13 shows. Overall, article realizations are less frequent after anchor words. In the first time slot, nearly all articles occur outside the context of anchor words. In the second time slot their relative occurrence is still only 6·2%, compared to 19·2% after other words. Only by the third time slot is their relative frequency (34·1%) on a par with that of article realizations after other words (37·3%). It seems then that,

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

contrary to what we have seen in the case of schwa fillers, the presence of anchor words actually inhibits the production of articles at first; only by the third time slot do the latter catch up.

It turns out that the phonological and lexical environments have a different impact on the production of articles than on the production of schwa fillers. Both are less frequent in rhythmically non-optimal environments (i.e. next to another weak syllable). But, whereas schwa fillers are extremely rare in utterance-initial position over the entire period of time, there is only a mild dispreference for full-fledged articles at the beginnings of utterances. Another clear difference is that schwa fillers predominate after anchor words, while the emergence of article realizations in this context seems inhibited until the third time slot. This delay may be explained by the inhibitory effect of the highly frequent sequences of anchor word plus schwa filler in the child's own speech.

DISCUSSION

Our analysis of Cato's fillers constitutes the first systematic study of Dutch fillers. In contrast to an earlier paper by Wijnen *et al.* (1994), who also considered fillers in their study on determiner omissions, we have chosen to widen the scope and to define fillers in purely phonological terms, regardless of phrasal position. This approach allows us to disentangle issues such as the relation between fillers and determiners, which are conflated if syntactic criteria are allowed into the definition. The resulting picture turns out to be more complex, but, in our view, rather more interesting.

On the one hand, our results provide partial confirmation of Wijnen *et al.*'s intuition that Dutch fillers are connected to articles. The distribution of schwa fillers in Cato's speech is skewed towards the distribution of articles in adult speech. For nasal fillers, this is only the case for a specific subgroup of nasal fillers, those in front of words starting with [h]. For both cases, however, the match between articles and fillers is far from perfect. Schwa fillers appear in contexts that do not allow articles in adult Dutch: outside NPs, in front of proper names, and before other determiners. Nasals before [h]-initial words not only appear before nouns beginning with [h], but they (temporarily) also appear before other words beginning with [h]. Because of these deviations from adult Dutch, we have rejected the hypothesis that those fillers are mere mispronunciations of adult-like articles. Instead, we want to describe them as proto-articles, elements that reveal an immature and even partially ungrammatical understanding of the distribution and shape of articles in the input (cf. Peters, 2001*a*). In contrast with these proto-articles, the child's full-fledged article realizations obey the adult restrictions much better: they rarely occur outside NPs or in

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

front of proper nouns. The only deviations concern a number of insertions of the singular article *een* in front of a plural noun, or before another determiner.

Wijnen *et al.* found that determiners, including both articles and fillers in determiner position, are affected by rhythmic and positional constraints: they are avoided in utterance-initial positions (by both children in their study) and utterance-internally before or after a weak syllable (by only one of the two children). Our approach distinguishes between nasal fillers, schwa fillers and articles, and we have found that these categorial distinctions are important. First of all, rhythmic and positional constraints only have a substantial effect on schwa fillers and articles. Nasal fillers seem merely conditioned by the first phoneme of the next word: they emerge before stops, and they appear before words (mainly nouns) beginning with [h]. Schwa fillers follow the predictions made by Wijnen *et al.* quite well: they are rare in utterance-initial position, and less frequent next to weak syllables. Articles are impacted by rhythm in much the same way as schwa fillers: they are more frequent when surrounded by stressed syllables, but whereas fillers hardly occur in utterance-initial position, this positional constraint is less strong for articles.

Whereas the link between fillers and articles, and various rhythmic and positional constraints are well-established in the literature, we have identified a third major determinant for Cato's filler productions, for which only scant mention can be found elsewhere: Cato's schwa fillers appear to be strongly conditioned by their lexical context. They are especially frequent after anchor words such as *is* '*is*', *ook* '*also*', *in* '*in*', *op* '*on*', *gaat* '*goes*' and *niet* '*not*'. An analysis of Dutch child-directed speech supports the hypothesis that these co-occurrences may be based on the distribution of articles in the input: articles are frequently preceded by *is, ook, in, op, gaat* or *niet*.

What do these observations tell us about children's incipient understanding of the ambient language? First of all, they give us some insight into the acquisition of articles in Dutch. Cato did not acquire this grammatical category in an instantaneous way, but passed through an intermediate stage with proto-articles, which revealed an incomplete understanding of the form and distribution of adult articles. Similar findings have been reported for other languages. In the introduction, we mentioned the example of a French child who produced proto-articles, which gradually approximated the forms and distribution of adult determiners, after having passed through a stage in which these same elements were merely inserted for rhythmic purposes (Veneziano & Sinclair, 2000). But while the French child initially focused on the rhythmic qualities of the ambient language, our Dutch child focused on surface characteristics of a different type: she discovered some highly

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

frequent co-occurrence patterns of words and articles in the ambient speech, and let them determine her own realizations of proto-articles.

As other studies have already noted, the mere existence of proto-articles goes against (generativist) theories of article acquisition, 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 (a.o. Valian, 1986). If this were true, the child would have no need for article-like elements, which, while clearly modeled after the adult targets, display different distributional restrictions.

But there is more. Traditionally, the claim about an early adult-like determiner category has been supported by the observation that errors in the use of articles are (nearly) absent in young children's speech (a.o. Abu-Akel, Bailey & Thum, 2004; Valian, 1986). For instance, Valian (1986) found that English-speaking children frequently omit determiners, signaling that the child still has to learn when the determiner slot may remain empty. But if a determiner was used, it was in the appropriate context: in front of a noun or an adjective. Furthermore, articles were found to be mutually exclusive with other determiners: the English children in Valian's study did not produce sequences like *a that cookie*. Our study, however, demonstrates that proto-articles not only occur in non-adult-like contexts, but that children's usage of full-fledged articles is not necessarily error-free either. Sequences of two determiners do occur in our data. Moreover, Cato often uses a singular article in front of a plural noun. Two studies of English-speaking children (Pine & Lieven, 1997; Pine & Martindale, 1996) prove that our data are not exceptional in this respect. Like Cato, their subjects produced singular determiners before plural nouns, and sequences of determiners. Moreover, they produced a number of utterances in which articles and other determiners are not followed by nouns or adjective+noun sequences. Such errors are rare in Cato's realizations of full-fledged articles, but occur with her schwa fillers.

Another correspondence between our study and the English studies by Pine and colleagues is that the English-speaking children were guided by lexical patterns as well. They often used articles in contexts as 'there's a+X', 'want a+X', 'a+X', which frequently occur in the ambient speech. A few such lexical contexts could account for more than half of the realizations of articles in their speech. Our own study confirms the importance of lexical context, and situates its impact in the precursors of articles. Moreover, we offer a viable explanation for these lexical effects: the co-occurrences of words and schwa fillers in the child's speech appear to be based on distributional characteristics of articles in the input.

This brings us to the central observation of this study. Our data support the claim that a child's early mastery of the ambient language can be guided by its distributional characteristics through a mechanism of distributional

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

learning. This mechanism shows up in Cato's usage of schwa fillers: once she has grasped the frequent occurrence of short words with a schwa (articles) in the input, and discovered that these words often occur in the same lexical contexts, she mimics these distributional characteristics by inserting phonologically similar schwa fillers in the same contexts. This child's sensitivity to distributional regularities also manifests itself in her usage of nasal fillers before words starting with [h], albeit on the phonological level. While nasal fillers originally occur only in front of nouns beginning with [h], probably in imitation of sequences of *een* + noun in the input, the child starts to generalize them to similar phonological environments outside the original grammatical context during the second time slot. This generalization is based on distributional learning of the phonological regularity in her previous utterances.

To our knowledge, this is the first study of fillers that reveals such a strong impact of distributional learning. Distributional learning itself is well-known as a powerful learning mechanism, which functions with visual, as well as auditory, or speech stimuli from early on (a.o. Kirkham, Slemmer & Johnson, 2002; Saffran, Newport & Aslin, 1996). From the fact that one-year-old children are able to learn adjacent dependencies of words in strings during listening experiments (a.o. Gómez & Gerken, 1999), it logically follows that these dependencies will also shape the distribution of words in their own speech, and this is exactly what we have found. Our results thus contribute to the growing body of evidence that children's early speech production is strongly affected by distributional learning. Several studies have reported that function words and inflections emerge earlier in some word contexts than in others ('low-scope' or 'item-based' learning, a.o. Tomasello, 2000). Often the choice of these word contexts appears to be influenced by frequencies of co-occurrence in the input (a.o. Theakston, Lieven, Pine & Rowland, 2002).

What our study adds to this evidence is the observation that the impact of distributional learning does not necessarily apply to all phonological realizations of the same target to the same extent or in the same way. Apart from the nasal fillers that occur in the limited context of nouns starting with [h], this child produces two types of attempts at articles, schwa fillers and full-fledged articles: the articles emerge while the schwa fillers are still frequent. But in spite of their co-occurrence, only the distribution of schwa fillers mirrors highly frequent sequences of words and articles in the input. In contrast to her schwa fillers, Cato's full-fledged articles are less rather than more frequent after words that often precede articles in the input. Instead, her full-fledged articles reveal a more advanced understanding of the restrictions of article usage in the ambient language than the concurrently produced schwa fillers: already during the second time slot, she only inserts articles at the beginnings of NPs. This knowledge

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

is not transferred to the schwa fillers, which she continues to produce outside NPs. We must conclude from this that although schwa fillers and articles are both modeled after articles in the input, they constitute distinct categories with different phonological representations and different distributional features. Moreover, these categories have their own way of dealing with rhythmic and positional constraints. Both schwa fillers and full-fledged articles are avoided next to a weak syllable. But the constraint against utterance-initial position has a stronger impact on the schwa fillers than on the article realizations. The child's better understanding of the adult distribution of articles as is reflected in her use of articles probably helps her to override this positional constraint to some extent.

Cases where different phonological representations of the same target category constitute distinct categories are not noted often in the literature, probably because they require the investigator's attention to the phonological form of a child's utterances (Johnson, 2000). In many child language databases this is a priori impossible, since a phonological transcription is absent. One of the merits of this study is that we have selected a database with phonological transcription, and used this information to a maximal extent: during the annotation of fillers and articles, we relied on the phonological forms only, and abstracted away from the grammatical position in which they occur. This methodology contrasts with the methodology applied by Wijnen *et al*. (1994): they abstracted away from the phonological forms, and instead focused on the grammatical position. As a consequence, we do not know whether their subjects applied the same distinctions between phonological forms as ours. Moreover, it remains unclear whether this is typical for Dutch children or not.

This case study gives rise to many other intriguing questions about the extent of inter-individual variation within and across languages. These questions concern the impact of lexical co-occurrence patterns as well as the scope of the phonological and morphosyntactic constraints reported in this study. Future studies should also compare the relative weights of the phonological, lexical and morphosyntactic factors in other children's filler insertions, and their evolution over time. In the Introduction we mentioned two different scenarios. In the first scenario, children start with a purely rhythmic stage of filler insertions and progress to a morphosyntactic stage. In the second, fillers are essentially morphosyntactic in nature from the very start, and rhythm only plays a role as a secondary constraint. Neither one fits our data. Against the first scenario counts the finding that we did not observe an early stage in which rhythmic and positional constraints were the only or even most important determinants. The second scenario provides a better match with our observations, although it neglects the early and significant impact of lexical context. Moreover, it cannot account for the fact that a particular subset of fillers, in our case the nasal fillers, was

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

(almost) entirely conditioned by phonemic environment. It remains to be investigated whether this specific scenario fits other children's development, and to what extent it is determined by the properties of the ambient speech.

REFERENCES

Abu-Akel, A., Bailey, A. L. & Thum, Y. M. (2004). Describing the acquisition of determiners in English: A growth modeling approach. *Journal of Psycholinguistic Research* **33**, 407–24.

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*, 1207–20. Somerville, MA: Cascadilla Press.

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.

Cameron-Faulkner, T., Lieven, E. & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science* **27**, 843–73.

Dabrowska, E. (2001). Discriminating between constructivist and nativist positions: Fillers as evidence of generalization. *Journal of Child Language* **28**, 243–5.

Feldman, A. & Menn, L. (2003). Up close and personal: A case study of the development of three English fillers. *Journal of Child Language* **30**, 735–68.

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.

Gómez, R. L. & Gerken, L. A. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition* **70**, 109–35.

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

Johnson, C. E. (2000). What you see is what you get: The importance of transcription for interpreting children's morphosyntactic development. In L. Menn & N. Bernstein Ratner (eds), *Methods for studying language production*, 181–204. Hillsdale, NJ: Lawrence Erlbaum.

Kilani-Schoch, M. & Dressler, W. U. (2001). Filler+infinitive and pre- and proto-morphology demarcation in a French acquisition corpus. *Journal of Psycholinguistic Research* **30**, 653–85.

Kirkham, N. Z., Slemmer, J. A. & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition* **83**, B35–B42.

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

Lleo, C. (2001*a*). Early fillers: Undoubtedly more than phonological stuffing. *Journal of Child Language* **28**, 262–5.

Lleo, C. (2001*b*). The transition from prenominal fillers to articles in Spanish and German. 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*, 713–37. Somerville, MA: Cascadilla Press.

Lopez-Ornat, S. (2001). Fillers: How much do they generalize? *Journal of Child Language* **28**, 243–5.

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**, 575–86.

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press

Peters, A. M. (1997). Language typology, prosody, and the acquisition of grammatical morphemes. In D. Slobin (ed.), *The crosslinguistic study of language acquisition*, 136–97. Hillsdale, New Jersey: Lawrence Erlbaum.
Peters, A. M. (2001*a*). Filler syllables: What is their status in emerging grammar? *Journal of Child Language* **28**, 229–42.
Peters, A. M. (2001*b*). 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*, 121–56. Amsterdam/Philadelphia: Benjamins.
Peters, A. M. & Menn, L. (1993). False starts and filler syllables: Ways to learn grammatical morphemes. *Language* **69**, 742–77.
Pine, J. M. & Lieven, E. V. M. (1997). Slot and frame patterns and the development of the determiner category. *Applied Psycholinguistics* **18**, 123–38.
Pine, J. M. & Martindale, H. (1996). Syntactic categories in the speech of young children: The case of the determiner. *Journal of Child Language* **23**, 369–95.
Saffran, J. R., Newport, E. L. & Aslin, R. N. (1996). Statistical learning by 8-month-old infants. *Science* **274**, 1926–8.
Theakston, A. L., Lieven, E. V. M., Pine, J. M. & Rowland, C. F. (2002). Going, going, gone: The acquisition of the verb 'go'. *Journal of Child Language* **29**, 783–811.
Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition* **74**, 209–53.
Valian, V. (1986). Syntactic categories in the speech of young children. *Developmental Psychology* **22**, 562–79.
Van Kampen, N. J. (1994). The learnability of the left branch condition. In R. Bok-Bennema & C. Cremers (eds), *Linguistics in the Netherlands* **199**, 83–94. 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.
Vihman, M. M. (1996). *Phonological development: The origins of language in the child*. Oxford: Blackwell.
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. & Bol, G. (1993). The escape from the optional infinitive stage. In A. de Boer, J. de Jong & R. Landeweerd (eds), *Language and cognition* **3**, 239–48. Groningen: University of Groningen.
Wijnen, F., Krikhaar, E. & den Os, E. (1994). The (non)realization of unstressed elements in children's utterances: Evidence for a rhythmic constraint. *Journal of Child Language* **21**, 59–83.

https://doi.org/10.1017/S0305000908008982 Published online by Cambridge University Press