

Root infinitives in Dutch early child language: an effect of input?*

FRANK WIJNEN

Utrecht University

MASJA KEMPEN AND STEVEN GILLIS

University of Antwerp UIA

(Received 2 February 1999. Revised 2 November 2000)

ABSTRACT

Children who acquire Dutch as their first language show a strong preference for using infinitival verb forms during the early stages of grammatical development. This exemplifies the ‘root infinitive’ (RI) phenomenon, which has played a significant role in recent discussions on the development of syntax. Most accounts proposed thus far invoke an immaturity of the child’s grammatical competence. We explore the possibility that the early predominance of infinitival forms is related to patterns in the language input. We analysed a corpus of utterances addressed by two Dutch-speaking mothers to their two- to three-year old sons. Root infinitive utterances amounted to 10%, and auxiliary-plus infinitive main verb constructions, which in terms of word order are maximally similar to RIs, constituted 30% of all verb-containing utterances. These figures render an account in terms of exposure to utterance structures unlikely. There is a moderate (but significant) correlation between frequency of occurrence of individual verb forms in the input and age of acquisition. However, infinitive verb forms are often acquired earlier than their input frequency would predict, and this may be related to an overall increased salience, due to their systematic appearance in sentence-final position and their relatively high conceptual transparency as compared to finite verbs.

[*] The research reported in this paper was supported by grants of the Dutch-Flemish Co-operative Programme on Language and Culture (VNC, nrs. G.2201.96 and 200-41.031), sponsored by the Belgian National Fund for Scientific Research (FWO) and the Netherlands Organisation for Scientific Research (NWO). Address for correspondence: Frank Wijnen, Utrecht Institute of Linguistics OTS, Achter de Dom 24 (1.07), 3512 JP Utrecht, The Netherlands. e-mail: frank.wijnen@let.uu.nl

INTRODUCTION

Nobody will deny that natural language acquisition requires two resources: a mechanism on the part of the child capable of generating linguistic knowledge, and primary data, in the form of utterances directed to the language learner. The latter is generally known as ‘input’, the former may be called the ‘Language Acquisition Device’ (LAD). Stated quasi-formally, LAD is a function which, when applied to the primary data, yields a language: $L = \text{LAD}(\text{input})$. There are deep controversies over the nature of the LAD, which we would want to settle with the help of arguments that refer to L , the language generated by the learner. And indeed, patterns in child language are quite systematically ascribed to specific qualities of the LAD, on the assumption that input is a constant, conforming to what our intuitions, or possibly grammar books, tell us. The possibility exists, however, that the primary data may not be what intuition or normative grammar dictate. Conceivably, the acquisition process is more adequately represented by a formula like $L = \text{LAD}(\text{Filter}(\text{input}))$, in which ‘Filter’ refers to a process resulting in some elements of input utterances having a stronger impact on the LAD than others. Let us call these salient elements. Thus, the possibility arises that children’s immature language is in some way a function of what is salient, or, in other words, what is selected from the input.

Root infinitives in early grammar

From this perspective, we look at a well-documented phenomenon in early Dutch child language, *viz.* the occurrence of infinitival matrix verbs, so-called ‘root’ or ‘optional’ infinitives (Rizzi, 1994; Wexler, 1994). The acquisition of the finite–nonfinite contrast in Dutch, particularly its syntagmatic reflexes, has drawn quite a bit of attention over the last ten to fifteen years. Much of this research has been motivated by the relative complexity of verb placement (as compared to, e.g. English). Dutch is a VERB SECOND language (as are cognate languages such as German, Afrikaans and Danish). ‘Verb second’ refers to the constraint that the finite verb in independent clauses must be in a left-peripheral position, preceded by maximally one constituent, as illustrated in (1). There is another position for verbs, which is clause-final (disregarding some types of extraposition). In matrix clauses this position is accessible for nonfinite forms only (i.e. infinitives and participles, as well as verb particles, see examples 2), whereas in dependent clauses, both finite and nonfinite verbs must appear there (examples 3). These facts are potentially confusing to the language learner. Grasping the system depends on discriminating dependent from independent clauses, and finite from nonfinite verb forms.

- (1) a. *Maurits zoent Marilène.*
 Maurice kiss-3SG Marilyn
 'Maurice is kissing Marilyn'
- b. *Zoent Maurits Marilène?*
 kiss-3SG Maurice Marilyn
 'Is Maurice kissing Marilyn?'
- c. **Op hun trouwdag Maurits zoende Marilène.*
 on their wedding day Maurice kiss-SG-PAST Marilyn
- (2) a. *Maurits wil Marilène zoenen.*
 Maurice want-SG Marilyn kiss-INF
 'Maurice wants to kiss Marilyn'
- b. *Maurits heeft Marilène gezoend.*
 Maurice have-3SG Marilyn kiss-PPT
 'Maurice (has) kissed Marilyn.'
- (3) a. *Iedereen zag Maurits Marilène zoenen.*
 everyone see-SG-PAST Maurice Marilyn kiss-INF
 'Everyone saw Maurice kiss Marilyn.'
- b. *Ik zag dat Maurits Marilène zoende.*
 I see-SG-PAST that Maurice Marilyn kiss-SG-PAST
 'I saw that Maurice kissed Marilyn.'

Independent sentences with a matrix infinitive are highly marked, and acceptable only in special moods (exclamatives, exhortatives, jussives), or discourse contexts (e.g. elliptic answers to interrogatives). Against this backdrop, it is remarkable that Dutch children in the initial stage of syntactic development produce many utterances containing (what appears to be) an infinitive verb as its main predicate. These verbs usually have a *schwa* ending (corresponding to the standard pronunciation of the infinitive suffix /en/), and systematically occur in utterance-final position, as in examples (4).¹ Such 'root' or 'optional' infinitives are not unique to Dutch child language. Various studies have revealed the phenomenon in other languages, such as French, German, Russian, English and the Scandinavian languages (see Wexler, 1994).

- (4) a. *thee drinken* (Niek, 2;7)
 tea drink-INF
 'drink(ing) tea'
- b. *die helemaal kapot maken* (Niek 2;7)
 that + one altogether broken make-INF
 'smash that one'

[1] Next to these, forms similar to the imperative are often observed. In many cases, such imperatives appear to be fixed, unproductive routines such as *kijk es!* 'look-it!'. In our analyses, imperatives were not included, as their status with respect to the finite-infinite contrast is unclear.

- c. *ook ə paard stappen* (Peter 1;9.20)
 also horse step-INF
 ‘horse also trot(ting)’
- d. *mama radio aan doen* (Peter 2;0.7)
 mummy radio on do-INF
 ‘mummy switch on radio’

What appears to be particular for Dutch (and possibly German; Behrens, 1993), however, although we cannot be sure due to a paucity of detailed crosslinguistic longitudinal studies, is a marked developmental pattern. Recent quantitative longitudinal studies indicate that infinitives make up (nearly) 100 percent of the verbs during the first month (approximately) after the onset of combinatorial speech (Wijnen & Verrips, 1998). In other words, Dutch-speaking children do not begin to use finite verb forms (first/second position; null- or *t*-suffix) until after several weeks (or more) after the onset of multi-word speech. The first finite forms usually are modals and the copula/auxiliary *is*, supplemented by a limited set of stative lexical verbs. Since the infinitive verbs at this point almost without exception denote (transitive) actions, finite and nonfinite constructions are in complementary distribution (De Haan, 1987; Jordens, 1990; Wijnen, 1998). Later, when the child acquires the finite counterparts of the previously learned infinitives, and the periphrastic (Aux + Vinf) predicate construction, finite constructions begin to replace the root infinitives, and the frequency of occurrence of the latter drops quite rapidly.

Root infinitives have figured quite prominently in recent debates on syntactic development. Various hypotheses have been proposed to account for the observed facts (see Schönenberger, Pierce, Wexler & Wijnen, 1995). Each of these proposals assumes that root infinitives are the result of an immaturity of the child’s grammar, or, in other words, a restriction on the principles of UG that can be employed to map input data onto a language-specific grammar. Two general approaches can be discerned. The first, sometimes labelled ‘full competence’, represented by Rizzi (1994) and Wexler (1994, 1998), among others, holds that early syntactic competence is essentially identical to that of the adult, except for a single principle, due to which root infinitives are considered to be acceptable and well-formed in contexts where adults reject them.² The second approach, alternately

[2] Over the recent years, various proposals have seen the light, each of which assumes that the child’s grammatical competence is LACKING a certain principle, e.g. that the Complementiser Phrase (CP) must be the root of a derivation (Rizzi, 1994), or that the Tense feature must be specified (Wexler, 1994). Recently, however, Wexler has taken a different tack, by stating that the child’s grammar is more RESTRICTIVE than the adult’s: ‘The OI stage ... is the result of *one* extra, more severe constraint that UG has – the Unique Checking Constraint.’ (Wexler, 1998: 74). Despite the technical differences, the

referred to as ‘structure building’ or ‘incremental acquisition’, is most prominently represented in Radford’s 1990 book. The principal claim is that children’s early grammar is a ‘scaled-down’ version of the adult grammar, in that certain components or modules are lacking (or not yet operative), as a result of which particular syntactic representations cannot be generated, notably those that entail the functional categories of syntax (e.g. agreement, number, tense).

To explain developmental change, the grammar-based approaches exemplified here invoke the concept of maturation. As a result of a presumed ultimately biological process, the components or features of grammatical competence that were initially inaccessible or absent become available, so that previously unattainable representations enter the repertoire, resulting in language output that conforms to adult standards. The implementation of maturation as an explanatory principle varies in important details over the two approaches, but we will not linger on this, and concentrate on the predictions. On the full competence account, the early (optional infinitive) grammar comprises derivations that result in nonfinite structures. The transition from the optional infinitive grammar to the adult grammar is effectuated in a single step, which can be characterized as a shift towards a grammar that no longer has the ability to omit certain functional features or projections (Wexler, 1998). The developmental prediction is that a uniform optional infinitive stage – the initial stage – is followed by an ‘adult’ (or final) stage. This prediction would seem to be at variance with the available evidence for Dutch, since, to the extent that a stage in which finiteness is truly optional can be documented, it is preceded by a stage in which finite verbs do not occur (Wijnen & Verrips, 1998).

A structure building approach along the lines of Radford (1990) would be comfortable with the Dutch developmental facts, as it assumes that functional categories, which, among other things, are needed to construct finite sentences, are lacking from the child’s initial grammar. The maturational step here refers to the sudden burgeoning of the functional categories (which are all dependent on a single module of grammar), as a result of which most of the morphosyntactic processes and elements that were absent in the initial stage, arise at once and in concert. However, the structure building model runs astray in several areas. Children that acquire Italian, Spanish, or related pro-drop languages do not pass through an initial (optional) infinitive stage (Wexler, 1998). Similarly, Hamann & Plunkett (1998) have shown that Danish children’s early language does not display a predominance of infinitives. Rather, finite verbs seem to be favoured at first. As for Dutch, the structure building hypothesis would be embarrassed by the observation that

gist of these hypotheses is similar, and most of the predictions relevant for the present discussion have remained the same.

finiteness marking at first appears to be limited to specific classes of verbs. In conclusion, the structure building approach as well as the full competence approach would appear to have distinct problems with detailed developmental data.

In this study we follow a different track. Rather than assuming that the Dutch developmental pattern summarized above is a direct reflection of a (possibly maturing) grammatical competence, we hypothesize that it depends on the way children perceive and process the ambient language. We focus on the predominance of infinitives in Dutch-speaking children's early language, and explore the idea that this phenomenon is due to the fact that infinitives in child-addressed speech are more salient, i.e. easier to perceive and process than finite verbs. In the literature on the relationship between input and language development three factors that purportedly contribute to relative salience, and hence order of acquisition, figure prominently: FREQUENCY, DISTRIBUTION (position in the sentence) and CONCEPTUAL TRANSPARENCY. In the ensuing paragraphs we will briefly summarize some of the most pertinent observations regarding these factors, and evaluate whether they shed light on the early predominance of root infinitives in early Dutch child language.

The effects of input frequency

The picture that emerges from studies of the effect of input frequency on children's uptake of linguistic forms is somewhat confusing. Cases have been made for the position that frequency affects child language, as well as for the position that it does not. To some extent the inconsistency may arise from the fact that at least two dependent variables have been looked at: frequency of use in child language, and order of acquisition (or: first appearance). Additionally, the levels of aggregation at which frequency effects are analysed differ across studies. Some have looked at the exposure to particular constructions types, irrespective of lexical content, others take the word (and the context it appears in) as a unit of measurement.

Does the frequency with which parents use particular constructions have an impact on child language? Gathercole (1986) looked at 3–5-year-old American- and Scottish-English-speaking children's use of simple past and past participles, against the background of dialectal variation. The use of the past participle in Scottish English is more widespread and hence more frequent than in American English. The child data mirrored the difference. On average, Scottish children used past participles 75 times more often than the American children in a 6 hour sample of spontaneous speech. Similar evidence was reported by De Houwer (1997), who looked at the use of past participles in an English-Dutch bilingual girl between ages 2;7 and 3;4. De Houwer observed that the girl's past participles vastly outnumbered simple past tenses in her Dutch utterances, whereas in her English utterances the

pattern was reversed. This contrast appeared to correspond well with the difference between the input languages.

Other studies have looked at precedence relations in acquisition. Demuth (1989) reported that Sesotho-speaking children start to use full-blown passives (including the equivalent of the *by*-phrase) at approximately age 2;3, which is considerably earlier than English-speaking children do. Demuth relates this difference to the fact that passives are much more pervasive in Sesotho than in English. In this study, as well as in those that looked at children's frequency of use, it is unclear whether there is evidence for an unconfounded effect of frequency. After all, the comparisons are between different languages or dialects, and hence the differences in input frequency, as well as in the children's output, are most likely related to grammatical differences, i.e. the specifics of form-function mappings.

A more direct test of the frequency hypothesis is provided in a study by Stromswold (1995), in which the age of first use of subject and object *wh*-questions was correlated with the frequency of use by the target children's caregivers, within a single (American English) dialect. This did not yield a statistically reliable result. Neither was the correlation between the relative frequency of subject and object questions in parental input and the difference in age of first use of these two constructions across subjects significant. Stromswold concludes that the input frequency of particular construction types at most has a minor impact on the order of acquisition. In a similar vein, Snyder & Stromswold (1997) concluded that the ages of acquisition of double object datives (*John will give Mary the book*) and *to*-datives (*John will give the book to Mary*) were unaffected by frequency of parental use.

Whereas the effects at the level of sentence structure are equivocal at best, studies of lexical acquisition have produced more consistent results. And this may turn out to be relevant for syntactic acquisition, as we will see. Gillis & Verlinden (1988) investigated vocabulary acquisition in the Dutch-speaking boy Maarten between ages 0;11 and 1;11. Contrary to Gentner's (1982) claim, they found that the proportion of nouns (both types and tokens) was considerably higher than of verbs in child-addressed speech. The rate of acquisition of nouns was higher than of verbs and the authors ascribe this to the frequency difference. At the same time, they point at a number of distributional and pragmatic factors, that are likely to amplify the frequency effect. We will return to these later.

Huttenlocher, Haight, Bryk, Seltzer & Lyons (1991) have shown that the average frequency of content words in a sample of 11 mothers' child-directed speech correlates significantly with the mean age of first appearance in the children's output. Naigles & Hoff-Ginsberg (1998) report a significant correlation between frequency of parental usage and order of acquisition in a set of 25 commonly-used verbs. Pine, Lieven & Rowland (1998) applied Huttenlocher *et al.*'s method to address the question whether the order of

acquisition of tensed and non-tensed (infinitive) verbs forms could be ascribed to relative frequency in the input. They determined the frequencies of use of all tensed and untensed verbs in the child-directed utterances of 12 English-speaking mothers, and found that the relative frequencies for the 25 most frequent items were highly similar across subjects. Correlating the log average frequency with the average age of acquisition yielded a coefficient of 0.79, which was highly significant, and the authors conclude that frequency of occurrence is a major determinant of the order of acquisition of tensed and untensed verb forms.

The upshot seems to be that there are frequency of exposure effects on lexical acquisition (at least for the open classes, cf. Brown, 1973). Pine *et al.* suggest that such lexical patterns can explain the inappropriate use of non-finite (infinitive) verbs in English child language. If particular verbs happen to occur predominantly as infinitives, the child will start using this form first. Pine *et al.* argue that learners may at first fail to notice that verbs need to agree with their subjects, because the input comprises conflicting cues, for example infinitival complement clauses such as 'I saw *the dog run*'. Note that applying the same logic to Dutch entails a complication, since, different from English, tensed and untensed verbs have different positional privileges. Explaining the preponderance of infinitive constructions in early Dutch child language in terms of a frequency effect at the level of individual lexical items (verb forms) would entail the supposition that lexical items are learned and stored in association with their distributional characteristics (canonical position).³

Linguistic factors that contribute to salience

Prosody and distribution (position) have been repeatedly indicated as determinants of the salience of lexical items. There are even indications that these factors have a stronger impact than frequency on order of acquisition. At least, they appear to significantly modify purported frequency effect. Gillis & Verlinden (1988) point out that not only nouns were more frequent in language input, but that they also occurred much more often in sentence margins than verbs did. Also, most single word utterances addressed to the target-child consisted of a noun. Thus, in the corpus under investigation, nouns were more perceptually salient than verbs. Similarly, the data collected by Naigles & Hoff-Ginsberg (1998) indicate that frequency of occurrence IN SENTENCE-FINAL POSITION is predictive of acquisition.

A by now classical observation in this regard is that the frequency of yes-no questions in child-addressed speech is correlated with English-

[3] The issue of lexical vs. structure-based effects is analogous to what has been termed the 'grain problem' in exposure-based approaches to syntactic ambiguity resolution in sentence processing (see Mitchell, Cuetos, Corley & Brysbaert, 1995).

speaking children's use of auxiliaries (Newport, Gleitman & Gleitman, 1977; Furrow, Nelson & Benedict, 1979; Gleitman, Newport & Gleitman, 1984). The fact that yes-no questions are the primary factor, rather than just the number of auxiliaries, suggests that the effect is due to frequency in combination with a surface property (sentence-initial position) that facilitates perceptual processing and segmentation of auxiliary verbs (the auxiliary clarification hypothesis, Richards, 1990).

Many studies point at the developmental advantage of lexical elements that occur in utterance-final position in child-addressed language. Slobin (1973), of course, is a classic reference. More recently, Fernald & McRoberts (1993) have shown that one-year-olds are better able to recognize a recently learned word if it occurs in utterance-final position. Shady & Gerken (1999), in an experimental study with two-year olds, found that their subjects performed better in a picture identification task when the target words were presented in utterance-final position. As these authors point out, the facilitative effect of utterance-final position on children's language learning is most likely related to the prosodic highlighting associated with this position. Utterance-final words have larger-than-average durations, and are marked by pitch peaks, in child-addressed language even more so than in adult-to-adult speech (Cruttenden, 1994). Also, they have a boundary at their right side, which in dialogue quite often corresponds to a pause, or a change of turn.

Finally, there are numerous suggestions that conceptual 'transparency', or 'simplicity' may be a predictive factor in lexical acquisition. Words with referents that are relatively easy to perceive and/or conceptualize tend to be acquired earlier than words with more abstract or nonsalient referents. This principle has been supposed to underlie the noun predominance in early acquisition (Gentner, 1982, but see Gillis & Verlinden, 1988). Basic level terms in English and Dutch are generally acquired before superordinates (Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976; Van Loon-Vervoorn, 1989). With regard to verbs, the observation made by Schlichting (1996) for Dutch that action verbs are acquired before all other types of verbs would also seem to be connected to perceptual and conceptual transparency. Actions are delimited and salient in the perceptual world – they can be pointed at or demonstrated. Hence, it would seem that for action verbs the connection between word and concept is relatively easy to grasp. This seems to hold more generally for *EVENTIVE* verbs (verbs denoting temporally and/or spatially bounded changes involving one or more actors). By contrast, the meaning of *STATIVE* verbs (e.g. 'to love', 'to fit') is much less transparent, as a consequence of which their acquisition is expected to be delayed.

Can RI preponderance be related to salience?

Returning to the Dutch developmental data, a first possibility is that the RI predominance in early language is due to a frequent exposure to this

construction type. *A priori*, there are several reasons to doubt the plausibility of this account. Firstly, the studies cited above suggest that caregivers' frequency of use of sentence types is not a robust predictor of acquisition precedence. More importantly, in view of the markedness of RIs in Dutch it seems unlikely that caregivers produce these constructions at a more than minimal rate. However, although the prediction is dubious, it is worthwhile to check the data, if only to be true to our own adage that general language descriptions should not be taken for granted where primary data to the language learner are concerned. Moreover, Lasser (1998) has claimed that adults produce RIs much more often than grammatical description would warrant, even outside of the restricted discourse contexts mentioned above.

If, in the end, RIs do not turn out to be very frequent in child-addressed language, perhaps another construction may do the trick, *viz.* the construction with a discontinuous predicate, in which a nonfinite, sentence-final lexical verb is the complement of a finite auxiliary in V2-position. It is conceivable that the child picks up on this construction, but fails to take it in and process it in its entirety. Notably, the auxiliary verb, which lacks referential meaning (semantic transparency), and is perceptually inconspicuous (because it is brief and unstressed), may be easily 'overlooked'. This explanation for the predominance of RIs in early child Dutch was originally proposed by Klein (1974). Ingram & Thompson (1996) have recently revived the idea by proposing that the typical modal interpretation of (German) RIs is the result of the fact that infinitives in the input most often occur as the complement of a modal auxiliary. Children are sensitive to the effect of the modal verb, but fail to process it as a separate lexical item, and consequently 'associate' modality with the infinitive form.

Under this hypothesis, we would expect to find that (modal) auxiliary-plus-infinitive constructions are much more numerous in child-addressed speech than utterances containing a simple (single-verb) finite predicate. This, in fact, is what Klein claims. Analysing the speech addressed by two mothers to their two-year old children during a one-hour recording, he found that the object-main verb order, which corresponds to the order of constituents in a root infinitive, occurred in over 70% of the mothers' utterances. The majority of these utterances contained auxiliary-plus-main verb predicates. Klein's data, however, are very limited, and so far, no-one has ever taken the trouble to verify them. Thus, the second goal of this study is to determine whether, indeed, Aux + V constructions outnumber simple finite constructions in child-addressed language.

Klein's suggestion appears to be based on the premise that frequency of occurrence at the level of sentence form is the effective factor. However, even if Aux + V constructions did predominate in the input, this does not tell us whether children pick up on this construction type (which would entail

that they are able to represent it in some form), or, as an alternative, that the putative effect on child language is mediated by individual verbs, as Pine *et al.* (1998) would hold. In order to check this, we need to correlate the frequencies of occurrence of individual verbs (verb forms) in the input with the order of appearance in child language. An interesting aspect of the lexical hypothesis is that it could, in principle, provide a straightforward explanation for the underrepresentation of finite forms in early child language. The idea is that if the set of finite forms in the input is restricted (many tokens represent the same types), relatively high frequency of use may promote early acquisition, but as the set of items to be acquired necessarily remains small, its applicability will be limited, and infrequent usage will ensue. Since, as Klein (1974) has claimed, many utterances in Dutch child-addressed language are of the Aux + V type, the chances are that the set of finite verbs may be small indeed. However, if the limitation is due to an overrepresentation of auxiliaries, and frequency is the main determinant of acquisition, it is unclear why Aux + V constructions would be absent from the child's early output.

If the variability of the set of infinitives is markedly greater than that of the finite verbs, this may contribute to a head start for infinitives in another way, since it would imply that the sentence-final position contributes more to the difference in meaning between sentences than the V2 position does. If we think of the child as a resource-limited language processor, we might expect that, in order to communicate efficiently, she would dedicate her limited capacity primarily to potentially high-information locations. The consequence may be that words that appear at those positions are acquired earlier and more easily.

The upshot is that we need to determine whether type and token frequencies of verb forms in language input can explain the order of acquisition. If they can, there is no need to take any other potentially salience-enhancing factors into account. But this may well turn out to be too rigorous. The research cited above strongly suggests that sentence-final position is a first-rate salience booster, and this is likely to contribute to the infinitive precedence in early Dutch child language. Conceptual transparency may also contribute. Schlichting (1996) has claimed that in informal adult discourse, the choice between the finite form and the infinitive of a lexical verb (embedded in a Aux + V predicate) is correlated with the semantic transitivity of the verb, as defined by Givón (1984). State verbs and ergative verbs show a strong tendency to surface as finite forms, whereas action verbs, particularly transitive action verbs, predominantly occur as infinitives. As it stands, nobody seems to have an explanation for this (statistical) pattern, and more importantly, there has so far not been an attempt to verify the claim, in particular for child-addressed language. If the suggested distribution turns out to hold, we may hypothesize that it facilitates acquiring infinitives,

TABLE 1. *Details of the sections of the corpora selected for analysis*

Phase	Matthijs				Peter			
	Age	MLUw	Utts. containing a verb		Age	MLUw	Utts. Containing a verb	
			Child	Mother			Child	Mother
One word	1;9.30-1;10.3	1.0	46	626	1;7.18-1;8.2	1.0	37	613
Early two-word	1;11.10-2;0.9	1.52	158	574	1;9.6-1;10.3	1.41	134	615
Optional infinitive	2;4.24-2;5.13	2.03	189	421	2;0.28-2;1.26	2.23	214	531
End	2;10.22-2;11.03	2.69	348	436	2;3.7-2;3.21	3.1	542	355
Total			741	2057			927	2114

Note: Mean Length of Utterance (words) calculations were done with the CLAN utility 'mlu'. Excluded from the count were utterances containing 'xxx' or 'yyy' codes, as well as the minors *ja* 'yes', *nee* 'no' and *oh* 'oh'.

since they can be more easily mapped onto the referential context than finite (lexical) verbs.

METHOD

For this study we used two longitudinal spontaneous speech corpora of the monolingual Dutch-speaking boys Peter and Matthijs. The corpora, which are available through CHILDES (MacWhinney, 1995), comprise series of fully transcribed recordings of natural, unstructured conversations between the target child, his mother (or occasionally father) and an investigator. Recordings were made once every fortnight, on average. On the basis of some fairly simple criteria, we divided the corpora into four sections, corresponding to typical phases in Dutch children's morpho-syntactic development (Wijnen & Verrips, 1998). (The term 'phase' is used descriptively.):

- (1) ONE-WORD PHASE: all utterances consist of one word.
- (2) EARLY TWO-WORD PHASE: the child has started to use multi-word utterances. Nearly all (lexical) verbs are infinitives, appearing at the end of sentences.
- (3) OPTIONAL INFINITIVE PHASE: root infinitives and finite sentences appear in roughly equal proportions.
- (4) END PHASE: root infinitives have almost disappeared from the child's output.

The proportions of root infinitives (i.e. utterances with a bare infinitive or participle as matrix verb) were derived from tallies of utterances that minimally comprised one verb. For each section two to three representative consecutive transcripts were selected, such that the number of child utterances was sufficient for quantitative analyses (see Table 1).

All utterances containing at least one verb were selected. From the sets of utterances corresponding to phases 2–4, we discarded one-word strings. The utterances identified for analysis were coded with respect to the positions of the verbs (if applicable) and the other major constituents (particularly, the subject and the object(s)). The coding for verbs also marked the distinction between finite and nonfinite morphology. These codes allowed us to (semi-)automatically count utterances containing various types of finite and nonfinite predicate structures. For the mothers' utterances, we also coded the addressee, which could either be the target child (Child-Directed Speech), an adult (Adult-Directed Speech), or unknown.

Lexical verbs in both the mothers' and the children's utterances were classified as either *EVENTIVE* or *STATIC*. Eventives are verbs that refer to dynamic changes that occur within a bounded interval, involving one or more actors. Transitive action verbs, which denote exchanges between an agent and a patient (e.g. 'to hit', 'to kiss') are the prototypical examples. Statives,

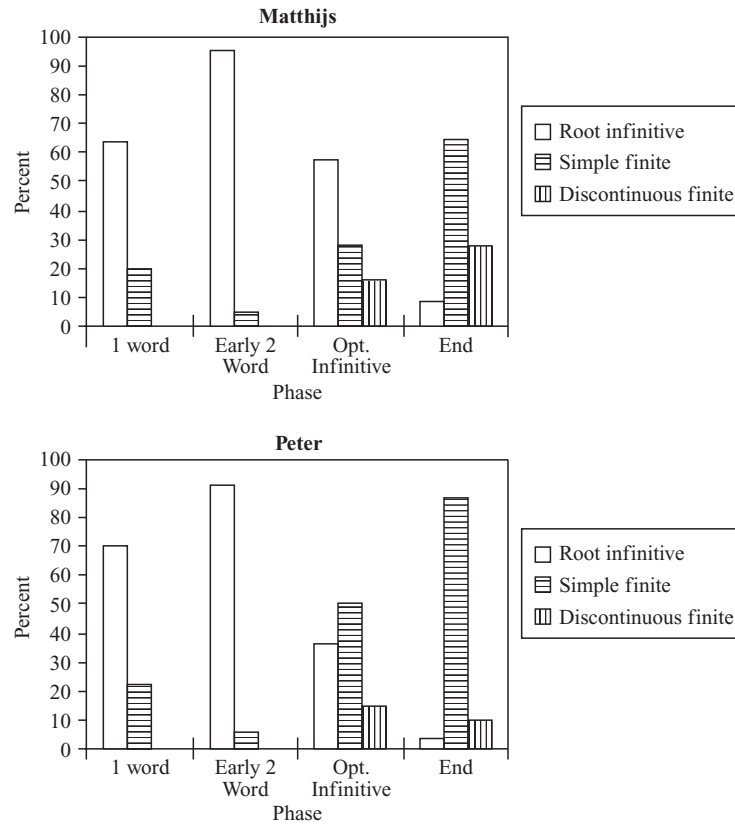


Fig. 1. Percentages of root infinitives, simple finite utterances, and utterances with a discontinuous predicate (Aux + V) as a function of developmental phase.

such as 'to love' or 'to know' denote relatively stable, unbounded conditions or situations. Besides meaning, various diagnostics for the classification are available (see Wijnen, 1998), the most simple of which is that (most) eventives can be substituted by 'to do', whereas statives can not.

RESULTS

The children

Figure 1 depicts the relative frequencies of three types of constructions in the two corpora, as a function of developmental phase. The raw numbers of utterances on which the percentages are based are given in Table 1. A small number of utterances (Matthijs: 13; Peter: 11) could not be classified, for

various reasons. These are not depicted, and hence the percentages per phase do not add up to 100. The data pertaining to infinitivals in fact subsume participial constructions, but these make up less than 5% of the relevant utterances, and do not occur in the first two phases. The graphs show that the token frequency of root infinitive constructions decreases over time, which is expected, as the four phases were defined on this measure. It is surprising, however, that the proportion of infinitival forms in the one word phase is lower than in the subsequent early two-word phase. Most likely, this is the effect of a slight difference in the composition of the sets of utterances for the one-word and early two-word phases. Starting with the early two-word stage, forms resembling imperatives were discarded from the analyses, as it is unclear whether they are finite or nonfinite. However, in one-word utterances imperatives cannot be distinguished from singular finite forms. Hence, all verb forms were included, and the label 'finite', should be taken to indicate: matching with a (singular) finite form. Figure 1 furthermore indicates that the growth rate of discontinuous predicates (Aux + V) is lower than that of simple finite predicates. This agrees with previously reported cross-sectional data (Bol, 1995).

Another way to look at the root infinitive phenomenon is to take stock of children's productive inventory of finite and nonfinite verb forms as it develops over time. This is relevant to the question whether infinitives are more easily acquired than finite verbs. Table 2 provides the relevant data. It gives an overview of the frequencies of verb forms, both by token and by type. Unclassifiable verb forms were excluded from these counts. Note that for both tokens and types, each morphological variant was counted as a separate item. So, for example *loop* 'walk-1SG' and *loopt* 'walk-2/3SG' are kept apart. Furthermore, the data are cumulative. This means that word forms that were first observed in phase X, were considered to be kept in the repertoire in the subsequent phases X+i. Thus, type counts are corrected for the recurrence of items in consecutive phases. No such correction was made for the token counts, since, naturally, this index must refer to the total number of times an item is uttered (within a circumscribed period). The number of Matthijs's tokens exceeds that of utterances in the first phase because word repetitions were counted as separate tokens, but not as separate utterances. Table 2 shows that infinitive types are, on the whole, more numerous than finite verb types, more strongly so in the early phases than in the later phases. We see that in the final developmental phase, the numbers of finite verbs and infinitives approach equality. The token frequencies show a similar pattern.

We conclude this section with a word on constituent order. Klein (1974) restricted his analyses to utterances that contained an object, and observed that the Object-Verb order was predominant. To verify this claim, we selected all two-word and longer utterances containing a constituent that, on

TABLE 2. *Cumulative numbers of finite and infinitive verb forms in Matthijs's and Peter's utterances, cross-classified over the four developmental phases*

Phase	Matthijs				Peter			
	Finite		Infinitive		Finite		Infinitive	
	Types	Tokens	Types	Tokens	Types	Tokens	Types	Tokens
One word	4 (20 %)	12 (24 %)	16 (80 %)	38 (76 %)	2 (33.3 %)	8 (24 %)	4 (66.7 %)	26 (76 %)
Early 2-word	9 (15.8 %)	23 (9.6 %)	48 (84.2 %)	217 (90.4 %)	9 (27.3 %)	21 (12.9 %)	24 (72.7 %)	142 (87.1 %)
OI phase	27 (27.6 %)	136 (26.7 %)	71 (72.4 %)	373 (73.3 %)	30 (42.9 %)	164 (37.6 %)	40 (57.1 %)	272 (62.4 %)
End	57 (39 %)	520 (51.5 %)	89 (61 %)	489 (48.5 %)	57 (50.9 %)	875 (69.3 %)	55 (49.1 %)	389 (30.7 %)

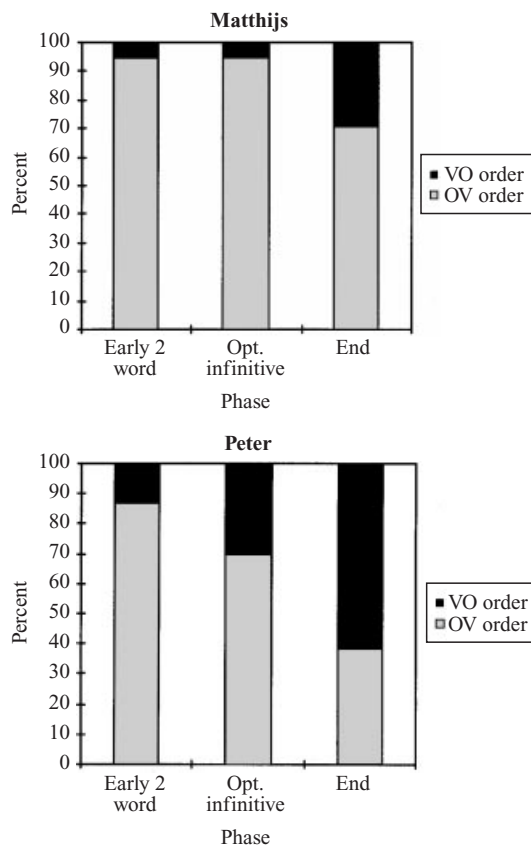


Fig. 2. Percentages of OV and VO orders in the children's multi-word utterances as a function of developmental phase.

semantic grounds, could be considered as either a direct object, an indirect object, or a prepositional object. 164 of such utterances were found in Matthijs's corpus, and 110 in Peter's. Examples are given under (4) below.

- (4) a. *mama radio aan doen* (Peter 2;0.7)
 mummy radio on do-INF
 'Mummy switch on radio'
 b. *eendje zien* (Matthijs 2;5.1)
 ducky see-INF
 '(I) look at the duck'

Figure 2 presents the relative frequencies of object-verb and verb-object order in the two corpora as a function of developmental phase. The classification depended on the number of verbs in the utterance. When an

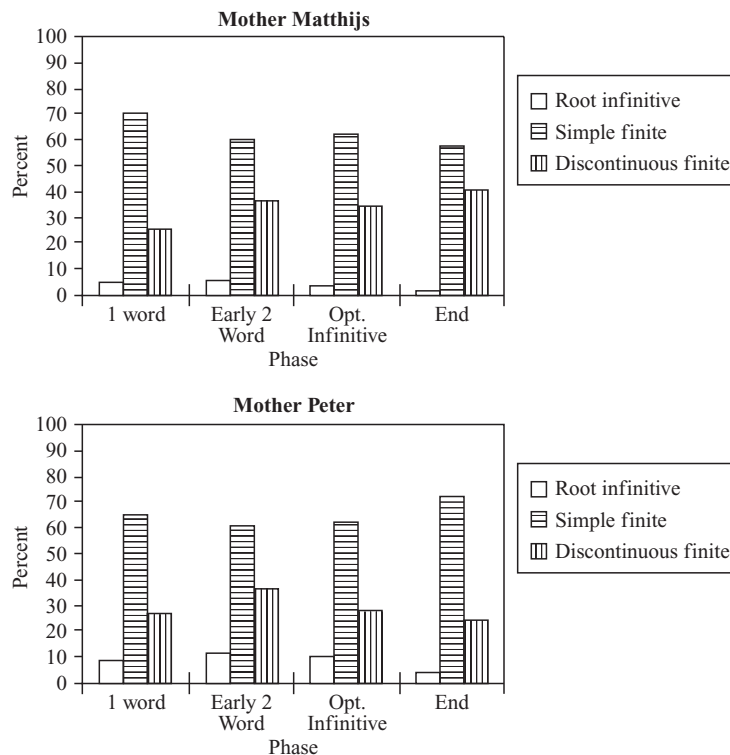


Fig. 3. Percentages of root infinitives, simple finite utterances, and utterances with a discontinuous predicate (Aux + V) in child-directed speech as a function of developmental phase.

utterance contained one verb (either finite or nonfinite) the position of the NP instantiating a theme, patient or goal was decisive, irrespective of whether a thematic role typical for the syntactic object could have been assigned by the verb present. For instance, in Dutch it is possible to have sentences with only a modal verb, e.g. *Ik mag een koekje* 'I may a cookie', which means 'I can have a cookie'. It is often assumed that modal verbs cannot assign a thematic role. Consequently, it is doubtful whether the apparent theme NP is the complement of the verb present. We ignored this complication and opted to classify sentences such as these as verb + object. When an utterance contained a (modal) auxiliary and a lexical verb however, the position of the theme/patient constituent *vis-à-vis* the lexical verb determined the classification.

Overall, the OV-order predominates. In 87–95 % of the early two-word phase utterances objects precede verbs. This percentage diminishes with age.

In the end phase, it has dropped to 71 % in Matthijs's corpus and even 38 % in Peter's. As observed before, the vast majority of all utterances in the early two-word phase are root infinitives. The same holds for the subset of utterances containing objects. These data confirm Klein's (1974) observation that the – grammatically correct – word order object-infinitive verb is predominant in early Dutch.

The mothers

Figure 3 is analogous to Figure 1, in that it shows the relative frequencies in CDS of root infinitives, simple finite utterances (containing a single finite verb), and discontinuous finite utterances (consisting of a finite auxiliary and a nonfinite lexical verb complement), as a function of the target-child's developmental phase. As expected, the mothers do not produce very many nonfinite sentences (10 % or less). About thirty percent of the utterances has a discontinuous finite predicate, and simple finite utterances make up the lion's share (around 70 %) of the corpora.

Of the child-addressed utterances produced by Matthijs' mother, 799 (38.8 %) contained either a direct, indirect, or prepositional object. For Peter's mother, the number was 795 (37.6 %). The relative frequencies of object-verb and verb-object order in these subsets of utterances are shown in Figure 4. The classification of utterances was based on the same rules as those employed for Figure 2. Figure 4 shows that the relative frequency of OV order varies around 60 %, without a clear indication of an age-related trend. Collapsed over the four phases, 62.4 % (314) of Matthijs's mother's and 56.8 % (288) of Peter's mother's OV utterances contain Aux+V predicates. Utterances with a single finite verb (mostly object topicalizations and *Wh*-questions: Wh/O-V_{FIN}-X) account for 21 % (105) of the OV-order utterances in Matthijs's mother's speech and 31 % (158) in Peter's mother's speech. The remainder are participial constructions and a few root infinitives.

Differences between left-peripheral and sentence-final verbs

In this section we will take a look at two factors that might contribute to the difference in salience between finite verbs and infinitives: variability and semantic transparency. By doing this, we are shifting our attention away from the effect of construction type frequencies on acquisition, to effects at the level of lexical items. In the next section we will push this a little further, by attempting to answer the question whether the precedence of infinitives in early Dutch is a function of input frequency on the level of individual verb forms.

The findings so far imply that many verbs that appear in verb-second position in child-directed utterances are auxiliaries. To verify this, we

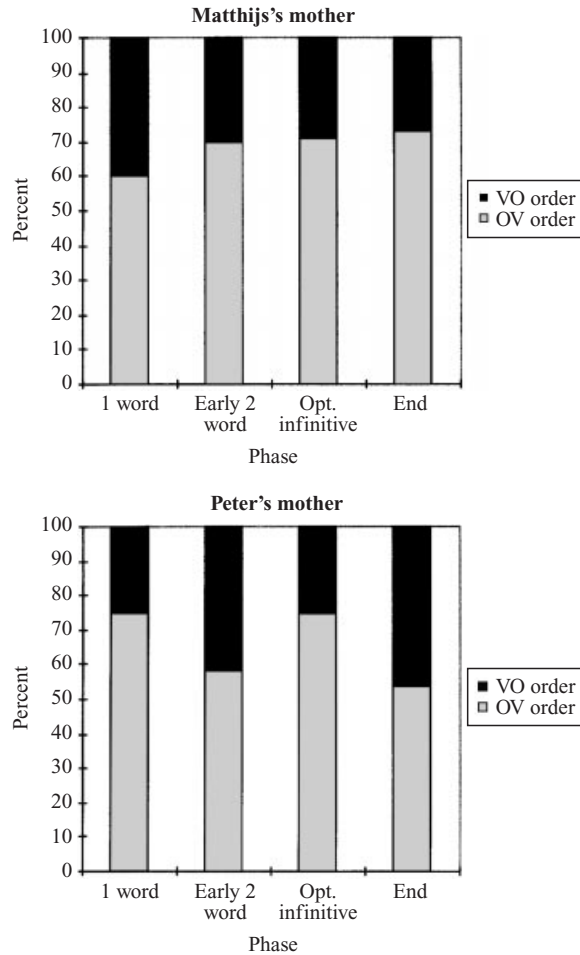


Fig. 4. Percentages of OV and VO orders in child-directed speech as a function of the target child's developmental phase.

classified the verbs in all the child-directed utterances according to position (verb second *vs.* final) and type (copula, auxiliary – including modals – or lexical). The resulting tallies are depicted in Figure 5. The graphs show an even distribution between (modal) auxiliaries/copulas and lexical verbs in verb second (left-peripheral) position, as opposed to an imbalance in final position, where almost all of the mothers' verbs are lexical. (Most of the few sentence-final auxiliaries and copulas appear in subordinate clauses, in which all verbs, finite as well as nonfinite, must be in final position.)

This finding might suggest that the number of different verb types in verb second position is lower than in sentence-final position. However, Table 3

ROOT INFINITIVES IN DUTCH EARLY CHILD LANGUAGE

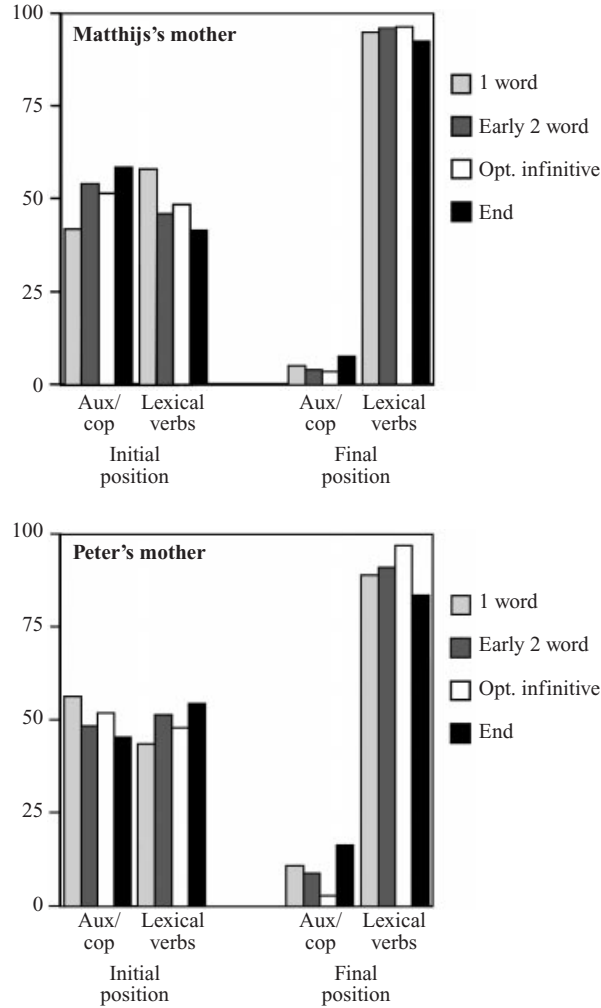


Fig. 5. Mothers' percentages of auxiliaries and lexical verbs in V2- and sentence-final positions, cross-classified with developmental phase.

shows that there is hardly a difference between the two positions in this respect. However, since a considerable number of left-peripheral finite verbs are selected from the restricted set of (modal) auxiliaries and copulas, the TYPE-TOKEN RATIO (TTR) of finite verbs is most likely lower than that of sentence-final infinitives. Figure 6, which presents the TTRs of left-peripheral and sentence-final verbs, confirms this expectation. On a low TTR (few types and/or many tokens), the likelihood of encountering the same verb in two arbitrarily selected utterances is higher than on a high

TABLE 3. *Numbers of verb types per sentence position in child-directed speech, cross-classified against developmental phase of the target child*

Phase	Position in sentence			
	Matthijs's mother		Peter's mother	
	Initial	Final	Initial	Final
One-word	58	40	30	24
Early 2-word	53	59	38	43
Opt. inf.	45	37	46	39
End	49	43	48	27

TTR. Consequently, sentence-final (infinitive) verbs contribute more often to a difference in meaning between two randomly selected sentences than verbs in second position, and hence are on average more informative than verbs in verb second position.

A further question is whether verbs appearing in either of the two positions differ in terms of conceptual transparency. In Table 4, we present the numbers of STATIVE and EVENTIVE verb types in the two positions, both in CDS in child language. Since position is irrelevant in the children's single-word utterances, the one-word phase was disregarded. The counts refer to LEXICAL verbs only. All auxiliaries and modals, as well as items ambiguous between an auxiliary and a lexical verb reading (e.g. *heeft* 'have-3S') were discarded. Transitive action verbs, a prototypical subclass of the eventives, are singled out. A limited number of verbs could not be classified, because, for instance, they were ambiguous between an eventive and a stative reading.

Table 4 shows that on the whole eventive verb types are more numerous in CDS than stative verbs. However, the extent to which eventives outnumber statives is different for the two positions. Whereas in verb second position the set of eventives is 3 to 4 times larger than the set of statives, it is 12 to 20 times larger in the sentence-final position. Thus, verbs in sentence-final position are significantly more often conceptually transparent than those in V2 position (Matthijs's mother: $\chi^2(1) = 14.63$, $p < 0.001$; Peter's mother: $\chi^2(1) = 4.02$, $p < 0.05$). In the children's repertoire the proportion of statives is slightly, but not significantly smaller than in the input. The distributional difference between statives and eventives roughly corresponds to that in CDS.

Lexical frequency and order of acquisition

The similarity of the stative-eventive distributions across sentence positions in the mothers and their children suggests a match between mothers' and the children's lexical inventories. To further explore this, we looked at the

ROOT INFINITIVES IN DUTCH EARLY CHILD LANGUAGE

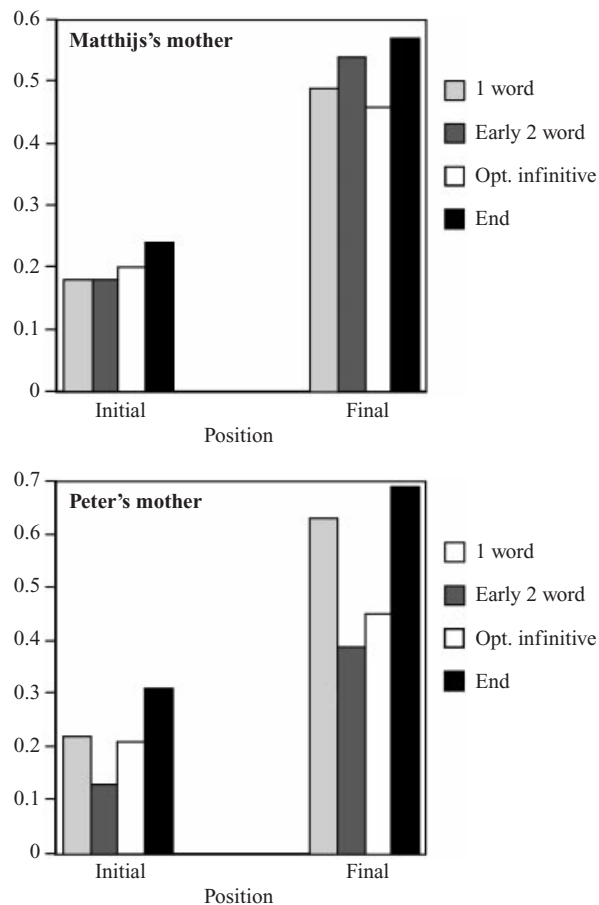


Fig. 6. Type-token ratios (TTRs) of verb forms in CDS as a function of position in the sentence.

TABLE 4. Numbers of verb types (stative, eventive) across sentence positions

	Sentence position	Stative verbs	Eventive verbs [transitive action]	Total	Unclassifiable
Mattheijs's mother	Initial (V2)	17	47 [13]	64	7
	Final	4	82 [40]	86	4
Mattheijs	Initial (V2)	3	11 [2]	14	3
	Final	1	36 [36]	37	2
Peter's mother	Initial (V2)	12	48 [15]	60	6
	Final	5	60 [32]	65	3
Peter	Initial (V2)	1	17 [5]	18	2
	Final	3	37 [15]	40	1

individual verbs that are used by the mothers and their children in the two verb positions. We address the question whether the expansion of the child's verbal inventory is related to the frequency with which particular verbs occur in specific positions in CDS. If input frequency is a decisive factor, we make two predictions: (1) The frequency with which a particular verb is used determines the time of acquisition. (2) The position/form in which a particular verb (lemma) is first used by the child is determined by the relative frequencies of occurrence of this verb in the two positions in the mother's utterances. In short, if a mother predominantly uses a certain verb as an infinitive (in sentence-final position), the child will do so too.

To test these predictions we correlated frequency of input with age of first appearance, following the approach described by Huttenlocher *et al.* (1991) and Pine *et al.* (1998). First, we made an inventory (with the aid of the CLAN application *FREQ*) of the hundred most frequent verbs in ALL OF the utterances recorded from the mothers, i.e. child-directed as well as other-directed speech. Each different verb form was treated as a separate lexical item. The raw frequencies of occurrence were converted into percentages, indicating relative frequency in the total set of items, which were in turn converted into base-10 logarithms. For each of the verb forms, age of first appearance was transformed into a rank number. If a form did not occur in child speech, it was assigned an arbitrary age-of-acquisition beyond the range studied. This occurred in three cases in each of the two sets of data.

If input frequency has a strong effect on acquisition order, we would expect to find negative correlations between log frequency and rank number. For the complete sets of verbs, $r = -0.30$ for Matthijs ($df = 98$, $p < 0.005$), and $r = -0.53$ for Peter ($df = 98$, $p < 0.001$). Thus, we find a moderate, but nonetheless significant correlation between input frequency and age of acquisition (first appearance). Scatterplots of the relation between input frequency and acquisition order are given in Figure 7. The sample reported on here contains some past participles. When these are removed, 92 cases remain for Matthijs and 95 for Peter. The correlations calculated over this restricted sample are slightly lower than those derived from the complete set but still significant, Matthijs: $r = -0.27$ ($df = 90$, $p < 0.01$), Peter: $r = -0.52$ ($df = 93$, $p < 0.001$).

The correlations presented so far were based on 'blind' counts of lexical forms only. Consequently, we were unable to distinguish between infinitives and (finite) plurals, both of which are marked by an *-en* suffix. This introduces some inaccuracy. Moreover, we looked at all recorded utterances of the mothers, including those that were not child-addressed. To evaluate the effect of these inaccuracies, we re-calculated the input frequencies of verb forms on the basis of the data from the four samples described before. The codes in these subsets allowed us to distinguish between child- and other-addressed utterances, and we could tell finite plurals apart from infinitives.

ROOT INFINITIVES IN DUTCH EARLY CHILD LANGUAGE

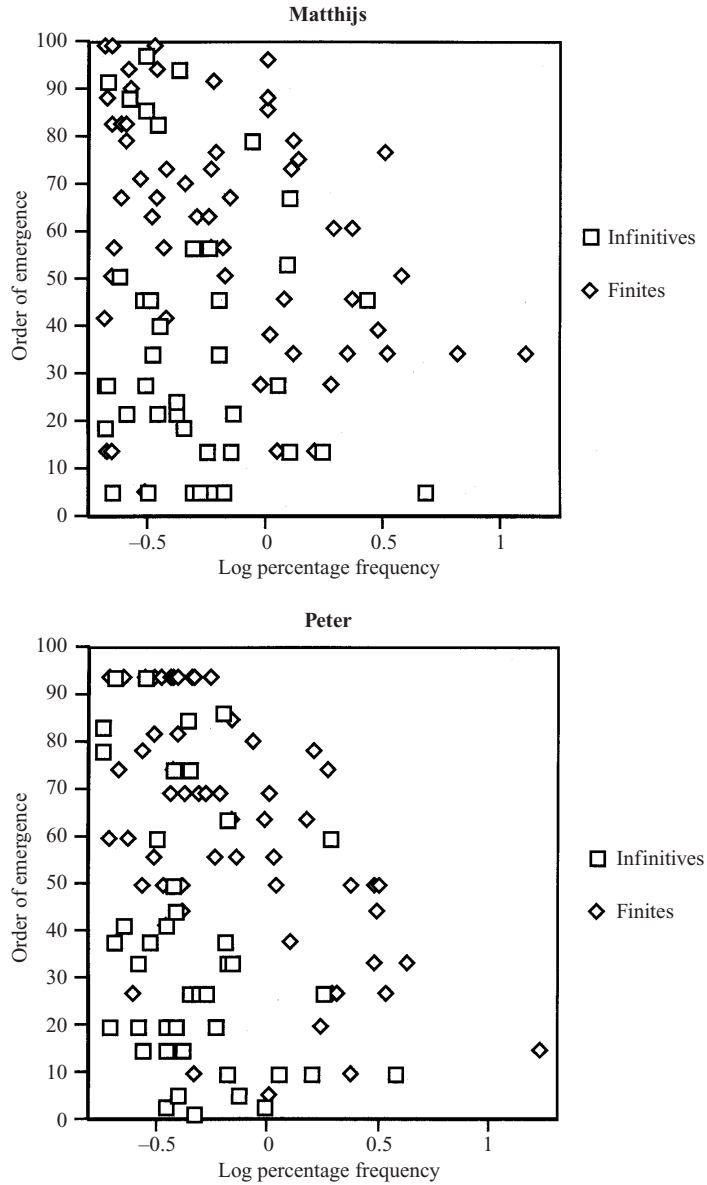


Fig. 7. Scatterplots of log relative frequency of verb forms in the input and rank order of first appearance in the children's language.

A disadvantage is, of course, that the numbers of lexical items in this analysis are considerably smaller than that for the original full corpus-based analysis (full corpus: 12549 words in Matthijs's mother's utterances, 2359 in samples; for Peter's mother: 8726 and 2388, respectively).

Based on counts in the fully analysed samples from the corpora, it can be estimated that at least 75 % of the mothers' utterances are child-addressed. Thus, if other-directed speech is different from CDS, it can only have had a relatively minor impact on the data reported in the previous paragraphs. The correlations between frequency of use in child-directed speech from the four corpus samples and acquisition rank order (as determined in the complete corpora!) are as follows, Matthijs: $r = -0.16$ ($df = 98$, $p < 0.2$), Peter: $r = -0.48$ ($df = 98$, $p < 0.001$).

It is somewhat disconcerting that in this last analysis, the correlation coefficient for Matthijs is considerably lower than that of the full-corpus analysis, and no longer statistically reliable. So far, we have no explanation for this. For Peter, the coefficients of the full corpus-based and sample-based analyses are in the same order of magnitude. In summary, it would seem that what have in hand is a moderate to weak statistical relation between input frequency and order of acquisition.

That other factors besides lexical frequency must be at stake is illustrated by two final observations. The first of these concerns the top-3 of most frequent verbs in the input and their ages of first use. Table 5 provides the

TABLE 5. *Top three most frequent verbs in input and the ages of first observed use by the child*

Rank	Item	Input freq. (%)	Age
(a) Matthijs			
1	<i>is</i> 'is-3S'	1635 (13.1 %)	2;1.7
2	<i>moet</i> 'must-SG'	837 (6.7 %)	2;1.7
3	<i>doen</i> 'do-INF'	605 (4.9 %)	1;11.24
(b) Peter			
1	<i>is</i> 'is-3S'	1486 (17.1 %)	1;9.20
2	<i>ga</i> 'go-1/2S'	372 (4.3 %)	1;11.3
3	<i>doen</i> 'do-INF'	331 (3.8 %)	1;9.6

relevant data. For both children, two out of the three topmost items are finite verbs, viz. the copula/auxiliary *is* 'is-3S', and the modal *moet* 'must-SG' and the auxiliary/main verb *ga* 'go-1/2S' for Matthijs and Peter respectively. The third item is an infinitive, *doen* 'do-INF'. What is striking is that whereas *is* is at least two times (up to five times) more frequent than the other items, it is not the one with the lowest age of first appearance in this list; that

is the infinitive 'catch-all' verb *doen*. More generally, finite forms of 'functional verbs' such as auxiliaries and copulas are predominant among the most frequently used verbs in the input speech. Eight out of the ten most frequent verbs in the speech of both Matthijs's and Peter's mothers are finite functional verbs. However, all of these items appear later than at least one infinitive.

The second observation concerns the order of acquisition of finite and infinitive forms of the same lexical verbs, in relation to their input frequency. In both corpora we identified the lexical verbs that occurred both as infinitives and finite forms in the children's speech ($n = 21$ for Matthijs, $n = 23$ for Peter). For each of these verbs we noted the order of first appearance of finite and infinitive forms as well as their frequencies in CDS. Table 6

TABLE 6. *Order of acquisition of finite and infinitive forms of lexical verbs, in relation to frequency in child-directed speech*

Order of acquisition	Frequency		
	fin > inf	fin < inf	fin = inf
		Matthijs	
Finite earlier	2	0	0
Infinitive earlier	13	5	0
Simultaneous	0	1	0
		Peter	
Finite earlier	3	1	0
Infinitive earlier	6	8	1
Simultaneous	3	1	0

shows that for most verbs in this overview (Matthijs: 18, Peter: 15), finite forms are acquired later than the infinitive. In a considerable number of these cases (13 out of 18 in Matthijs, 7 out of 15 in Peter) the infinitive does NOT occur more frequently in the input than the finite form. In Peter, three more finite forms and corresponding infinitives come in together, even though the finite form is more frequent. These data are indicative of the subordination of input frequency to other factors affecting intake, where infinitives are concerned.

DISCUSSION

The data we have reported show, once again, that children who acquire Dutch as their first language have a strong preference for using infinitival matrix verbs, during the early phases of linguistic development (Wijnen &

Verrips, 1998). The percentage of root infinitives is close to one hundred at the onset of the two-word stage, and diminishes over time, until it reaches the adult range (5–8%; Wijnen & Bol, 1993) around age 3. Our objective in this study was to answer the question whether the dominance of root infinitive sentences can be related to properties of the language input. Are there reasons to believe that Dutch infinitives are somehow more salient for the language learning child than finite verbs are?

First of all, following Klein's (1974) lead, we have considered the possibility that the root infinitive predominance is an input frequency effect at the level of syntactic constructions. It appeared, however, that root infinitives are rather infrequently used by caretakers ($\leq 10\%$), although their inclination to do so in child-addressed speech is notably stronger than when they talk to adults. Of the two mothers' adult-directed utterances in our corpora, less than 5% were root infinitives. Perhaps this difference is due to a tendency on the part of the mothers to adapt their language use to their children. Anyhow, there is no empirical basis for the claim that young Dutch-speaking children's abundant use of root infinitives is a direct reflection of patterns in the input.

Our next step was to look at the frequency of exposure to a construction that is maximally similar to the root infinitival: sentences with a periphrastic predicate consisting of a (finite) auxiliary and an infinitive main verb (Aux + V). The similarity is such that RIs might be considered as Aux + V constructions with a missing or elided auxiliary (Klein, 1974; Ingram & Thompson, 1996). It appeared that approximately 30% of the input utterances in our corpora were Aux + V constructions, which is not exactly an overwhelming majority. Constructions with an auxiliary, a sentence-final infinitive and an explicit (direct, indirect or prepositional) object make up 13–15% of all input utterances. This figure seems to be at odds with Klein's (1974) claim that Dutch children's early adherence to the OV_{INF} order is directly related to frequency of exposure. In fact, if Klein's argument were to hold up, the figures we have found would lead us to expect finite utterances (which most often have VO order) to take precedence in the acquisition of Dutch. Thus, our data do not support Klein's original proposal, but they do agree with the observations of Stromswold (1995) and Snyder & Stromswold (1997), which are unresponsive to an effect of input frequency at the level of construction types on the order of appearance.

The child data have made it clear that the root infinitive phenomenon in early Dutch is not just a matter of the child choosing to use the RI structure. The lexical inventories show that infinitive verbs (types) form the majority in our subjects' early verb lexicons. So it would seem that infinitive verb forms (qua lexical items) are more easily acquired than finite verb forms. Given that a considerable portion of the input utterances has the Aux + V format, and that Aux is a closed class, it might be conjectured that the 'backlog' of finite

forms in the children's lexical inventories is simply the result of a reduced number of items to be learned. (All of this based on the assumption that finite forms and infinitives, at least in the initial phases, are acquired as separate items.) However, our data show this to be incorrect. If lexico-semantic identity (i.e. the lemma) is taken as a criterion, there are no differences in the numbers of finite and infinitive verb types in the mothers' speech. When we discriminate between various inflections, the number of finite items (types) is even considerably larger than that of infinitives.

It would seem, then, that infinitives are indeed more easily acquired than finite forms. In order to relate this to input characteristics, we have to look for factors that contribute to salience at a lexical level. Most likely, sentence-final position is one such factor. The literature contains numerous suggestions that lexical items on the edges of utterances are easier for children to segment and identify than sentence-medial elements. Mostly, the impact of sentence-final position is ascribed to prosodic factors. Our results suggest an additional factor. We observed that the type-token ratio of sentence-final infinitives is considerably higher than that of finite verbs. This means that, all other factors being equal, sentence-final infinitive verbs can be expected to more often contribute to a difference in meaning between two arbitrary selected sentences than finite verbs do. In other words, sentence-final verbs are more informative than finite verbs. We surmise that this feature assists in drawing the child's attention to elements in sentence-final position, on the assumption that, generally speaking, it would seem to be adaptive for a capacity constrained processing system to focus on those elements in a stimulus array that are most informative.

The infinitives of Dutch CDS also have a conceptual advantage. Our data show that most infinitives (tokens) are lexical verbs, which are taken from a vocabulary that almost entirely consists of eventives, a large share of which denote transitive actions. By contrast, finite verbs in CDS very frequently are auxiliaries (which do not have much conceptual content to begin with). The repertoire from which the remaining lexical verbs are taken contains a considerable number of statives. Consequently, we may say that there is a difference in average semantic transparency between finite verbs and infinitives, in the sense that the latter tend to have referents that stand out more clearly in the perceptual world than the former. Newport *et al.* (1977) suggested that semantically transparent utterances have a stronger impact on the course of acquisition. Our data are compatible with this speculation.

We may, at this point, tentatively conclude that sentence-final position, high information load and semantic transparency make infinitives more easy to pick up and store than finite verb forms. Is there still a role to play for frequency of exposure, at the level of lexical items? Our analyses have yielded moderate correlations (at best) between frequency of verb forms in the input and age of first appearance. Thus, lexical (token) frequency does appear to

have a distinct, albeit limited effect. Some caution is called for, however, in view of the different outcomes of the various analyses, and, particularly, the inflatory effect of a small number of extremely frequent verb forms (such as *is*). To neutralize this latter effect, we looked at the order of acquisition of finite and infinitive forms of the same lexical verbs, in relation to their input frequency. Strikingly, in a majority of cases, the finite forms were more frequent than their infinitive counterparts. Nonetheless, they are, on the whole, acquired later.

It is remarkable that Pine *et al.* (1998), from whom we copied the correlational method, report a considerably higher correlation between input verb form frequency and age of acquisition (0.79). What springs to mind as an explanation for the difference between their and our results is the structural difference between English and Dutch. In English, finite as well as non-finite (root) verbs can and in fact often do occur in (the same) sentence-internal positions. Our data suggest that the main reason for a relatively low correlation between input frequency and age of acquisition in Dutch is that the factors contributing to the salience of infinitives mentioned above outweigh mere frequency effects. It is certain that the positional advantage of infinitives does not hold for English, and the eventive-infinitive association typical for Dutch is not expected to hold for English either (Hoekstra & Hyams, 1998). Being less affected by these variables, English-speaking children's acquisition of verb forms may be more strongly determined by mere frequency of exposure. That this conjecture is on the right track is can be verified by looking at the input frequency-age of appearance correlations for FINITE verbs only. In our data, these appear to be markedly higher than those for the total set of verbs (Matthijs: $r = -0.55$, $df = 56$, $p < 0.001$; Peter: $r = -0.72$, $df = 55$, $p < 0.001$). Thus, if we eliminate the variables position and semantic transparency, we see that the impact of frequency gains in strength.

To summarize, we think our data indicate that the early occurrence and abundance of root infinitivals in young Dutch-speaking children's speech is related to a conspiracy of factors that boost infinitives' salience for the language learner. A frequent exposure to utterance tokens that instantiate the root infinitive or auxiliary-plus-main verb construction type does not seem to play a part in this conspiracy. What does seem to be effective to some extent is input frequency of individual verb forms, but this appears to be significantly modulated by the effects of position and semantic transparency.

The recent literature has seen various attempts to explain young children's widespread use of root infinitives as the outcome of an immature grammar. In doing so, these grammar-based hypotheses take up the responsibility for the changes in linguistic behaviour we observe in children: whenever the output changes, this must be the reflection of a change inside the grammar. We argue that the early preponderance of root infinitives is a function of

input characteristics which affect the relative salience of these elements. In other words, Dutch children's early language patterns are determined by the 'filter' on the input. By implication, the behavioural development (decrease of RIs) is a function of a changing filter. It is clear that we do at this point not have a theory of how and why the filter changes. Be this as it may, the upshot of our proposal is that it places the 'motor' of linguistic development outside the grammar proper. We do not deny that children's grammars may change, but we think it unnecessary to ascribe such changes to structural alterations of the language faculty, biologically driven or not.

REFERENCES

- Behrens, H. (1993). *Temporal reference in German child language*. Ph.D. dissertation, University of Amsterdam.
- Bol, G. W. (1995). Implicational scaling in child language acquisition: the order of production of Dutch verb constructions. In M. Verrips & F. Wijnen (eds), *Papers from the Dutch-German Colloquium on Language Acquisition* (ASCLD 3). Amsterdam: Institute of General Linguistics.
- Brown, R. (1973). *A first language: the early stages*. Cambridge, MA: Harvard University Press.
- Cruttenden, A. (1994). Phonetic and prosodic aspects of Baby Talk. In C. Gallaway & B. J. Richards (eds), *Input and interaction in language acquisition*. Cambridge: C.U.P.
- De Haan, G. (1987). A theory-bound approach to the acquisition of verb placement. In G. de Haan & W. Zonneveld (eds), *Formal parameters of generative grammar. Yearbook 1987*. Utrecht: OTS.
- De Houwer, A. (1997). The role of input in the acquisition of past verb forms in English and Dutch: Evidence from a bilingual child. In E. V. Clark (ed.), *Proceedings of the 28th Child Language Research Forum*. Stanford, CA: CSLI.
- Demuth, K. (1989). Maturation and the acquisition of the Sesotho passive. *Language* **65**, 56-80.
- Fernald, A. & McRoberts, G. (1993, October). *Effects of prosody and word position on infants' lexical comprehension*. Paper presented at the Boston University Conference on Language Development.
- Furrow, D., Nelson, K. & Benedict, H. (1979). Mothers' speech to children and syntactic development: some simple relationships. *Journal of Child Language* **6**, 423-42.
- Gathercole, V. (1986). The acquisition of the present perfect: explaining difference in the speech of Scottish and American children. *Journal of Child Language* **13**, 537-60.
- Gentner, D. (1982). Why nouns are learned before verbs. In S. Kuczaj (ed.), *Language development, vol. 2: Language, thought, and culture*. Hillsdale, NJ: Erlbaum.
- Gillis, S. & Verlinden, A. (1988). Nouns and verbs in early lexical development: effects of input frequency? *Antwerp Papers in Linguistics* **54**.
- Givón, T. (1984). *Syntax, a functional-typological introduction*. (Vol. 1). Amsterdam: Benjamins.
- Gleitman, L., Newport, E. & Gleitman, H. (1984). The current status of the motherese hypothesis. *Journal of Child Language* **11**, 43-79.
- Hamann, C. & Plunkett, K. (1998). Subjectless sentences in child Danish. *Cognition* **69**, 35-72.
- Hoekstra, T. & Hyams, N. (1998). Aspects of root infinitives. *Lingua* **106**, 81-112.
- Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M. & Lyons, T. (1991). Early vocabulary growth: relation to language input and gender. *Developmental Psychology* **27**, 236-48.
- Ingram, D. & Thompson, W. (1996). Early syntactic acquisition in German: evidence for the modal hypothesis. *Language* **72**, 97-120.

- Jordens, P. (1990). The acquisition of verb placement in Dutch and German. *Linguistics* **28**, 1407–48.
- Klein, R. M. (1974). Word order: Dutch children and their mothers. *Publications of the Institute of General Linguistics* 9. University of Amsterdam.
- Lasser, I. (1998). *Finiteness in adult and child German*. Ph.D. dissertation, CUNY.
- MacWhinney, B. (1995). *The CHILDES project: tools for analyzing talk*. (2nd ed.). Hillsdale, NJ: Erlbaum.
- Mitchell, D., Cuetos, F., Corley, M. & Brysbaert, M. (1995). Exposure-based models of human parsing: evidence for the use of coarse-grained (non-lexical) statistical records. *Journal of Psycholinguistic Research* **24**, 469–88.
- Naigles, L. R. & Hoff-Ginsberg, E. (1998). Why are some verbs learned before other verbs? Effects of input frequency and structure on children's early verb use. *Journal of Child Language* **25**, 95–120.
- Newport, E., Gleitman, H. & Gleitman, L. (1977). Mother, I'd rather do it myself: some effects and non-effects of maternal speech style. In C. Snow & C. A. Ferguson (eds), *Talking to children: language input and interaction*. Cambridge: C.U.P.
- Pine, J., Lieven, E. V. M. & Rowland, C. F. (1998). Tense optionality and children's use of English verb morphology: testing the Optional Infinitive Hypothesis. Paper presented at the 1998 *Child Language Seminar*, September 4–6, University of Sheffield.
- Radford, A. (1990). *Syntactic theory and the acquisition of English syntax*. Oxford: Blackwell.
- Richards, B. (1990). *Language development and individual differences: a study of auxiliary verb learning*. Cambridge: C.U.P.
- Rizzi, L. (1994). Early null subjects and root null subjects. In T. Hoekstra & B. D. Schwartz (eds), *Language acquisition studies in generative grammar*. Amsterdam: Benjamins.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M. & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology* **8**, 382–439.
- Schlichting, J. E. P. T. (1996). *Discovering syntax. An empirical study in Dutch language acquisition*. Doctoral dissertation: University of Nijmegen.
- Schönenberger, M., Pierce, A., Wexler, K. & Wijnen, F. (1995). Accounts of root infinitives and the interpretation of root infinitives. *Geneva Generative Papers* **3**, 47–71.
- Shady, M. & Gerken, L. (1999). Grammatical and caregiver cues in early sentence comprehension. *Journal of Child Language* **26**, 163–76.
- Slobin, D. I. (1973). Cognitive prerequisites for the acquisition of grammar. In C. A. Ferguson & D. I. Slobin (eds), *Studies of child language development*. New York: Holt, Rinehart and Winston.
- Snyder, W. & Stromswold, K. (1997). The structure and acquisition of English dative constructions. *Linguistic Inquiry* **28**, 281–317.
- Stromswold, K. (1995). The acquisition of subject and object *Wh*-questions. *Language Acquisition* **4**, 5–48.
- Van Loon-Vervoorn, W. A. (1989). *Eigenschappen van basiswoorden*. Lisse: Swets and Zeitlinger.
- Wexler, K. (1994). Optional infinitives, head movement, and the economy of derivation in child grammar. In N. Hornstein & D. Lightfoot (eds), *Verb movement*. Cambridge: C.U.P.
- Wexler, K. (1998). Very early parameter setting and the unique checking constraint: a new explanation of the optional infinitive stage. *Lingua* **106**, 23–79.
- Wijnen, F. (1998). The temporal interpretation of Dutch children's root infinitives: the effect of eventivity. *First Language* **18**, 379–402.
- Wijnen, F. & Bol, G. W. (1993). The escape from the optional infinitive stage. In A. de Boer, J. de Jong & R. Landeweerd (eds), *Language and Cognition 3. Yearbook of the research group for theoretical and experimental linguistics of the University of Groningen*. Groningen: Centre for Language and Cognition.
- Wijnen, F. & Verrips, M. (1998). The acquisition of Dutch syntax. In S. Gillis & A. De Houwer (eds), *The acquisition of Dutch*. Amsterdam: Benjamins.