Developmental Science 10:5 (2007), pp 588–612 DOI: 10.1111/j.1467-7687.2006.00629.x

PAPER

Modeling continuity and discontinuity in utterance length: a quantitative approach to changes, transitions and intra-individual variability in early grammatical development

Dominique Bassano[1] and Paul van Geert[2]

1. *Laboratoire Structures Formelles du Langage, CNRS-Université Paris 8, France*
2. *The Heijmans Institute, University of Groningen, The Netherlands*

Abstract

The confluence of an anomaly such as a growth spurt or a temporary regression on the one hand and a temporary increase in intra-individual variability on the other hand, forms a strong indicator of a major transition in early language development. Data concern one-word (W1), two- and three-word (W2–3), and four-and-more-word (W4+) utterances from two French children during their second and third years. A dynamic growth model was fitted, based on a structure of supportive, conditional and competitive relationships. Using a statistical simulation method, we showed two striking peaks of variability in addition to a temporary regression or rapid growth in the proportions of W1, W2–3 and W4+ utterances. We argue that these phenomena show transitions corresponding to critical points in grammatical development, which could be indicative of the emergence of simple combinatorial and syntactic stages of language successively. Our results emphasize the relevance of time-serial data and of intra-individual variability in the study of developmental transitions in general.

Introduction

A main issue in developmental psychology is to determine whether and how new structures emerge and how change is shaped within and across the various fields of child development, among which the development of language is a central domain (Emde & Harmon, 1984; Elman, Bates, Johnson, Karmiloff-Smith, Parisi & Plunkett, 1996; MacWhinney, 1999; van Geert, 1991, 1994, 1998b). If the new structure is important enough to transform the characteristic features of performance, its emergence marks a developmental transition. Such transitions entail continuity as well as discontinuity and are often accompanied by qualitative indicators, such as changes in intra-individual variability (van der Maas & Molenaar, 1992; De Weerth, van Geert & Hoitink, 1999; van Dijk & van Geert, 2007), but also by rapid growth and temporary regressions (van Geert, 1998a).

This article is based on the analysis of longitudinal data from the free speech of two French children during their second and third years. Its aim is twofold. First, it introduces a quantitative approach to the analysis of data on utterance length, based on a combination of curve fitting, dynamic systems modeling and the statistical analysis of intra-individual variability. The second aim is to provide evidence of the occurrence of two discontinuities in the data. Although the emphasis lies on showing that such discontinuities occur, we will also suggest a possible explanation, namely that the discontinuities mark the transition from a holoprastic to a simple combinatorial mode and from a simple combinatorial to a more abstract syntactic mode of language production. The basic evidence for these transitions consists of a combination of, on the one hand, a dynamic systems model of the data, suggesting the occurrence of growth spurts (and eventually also a temporary regression) and, on the other hand, statistical modeling of the intra-individual variability of utterance length, which shows two sudden and temporal increases in variability.

The structure of this article is as follows. We begin with a discussion about utterance length and how it might relate to grammatical development. The section concludes with a prediction of two transitions in early grammatical development. The second section introduces dynamic systems theory as a theoretical framework for developmental transitions. The Data section describes utterance

Address for correspondence: Dominique Bassano, Centre Henri Piéron, 71 avenue Edouard Vaillant, 92774 Boulogne-Billancourt Cedex, France; e-mail: dominique.bassano@univ-paris5.fr or vangeert@inn.nl

length data collected from two French-speaking children. It is followed by descriptions of two types of data analysis. The first presents a dynamic growth model of the utterance length data, and the second presents an analysis of intra-individual variability in utterance length and provides evidence of two statistically significant peaks in variability. The article concludes that the results from the dynamic model and the analysis of variability patterns support the prediction of two major transitions.

Utterance length in early language acquisition

To what extent is utterance length an index of grammatical development?

More than 30 years ago, mean length of utterance (MLU, calculated in morphemes rather than in words) was promoted by Brown (1973) as 'an excellent simple index of grammatical development' that made it possible to match children on levels in language development. Brown linked growth in MLU to movement through five stages from MLU 1.75 to MLU 4.50. Although our present work focuses on the development of utterance types of different word lengths rather than on MLU (i.e. one-word, two-word utterances, etc.), questions raised about MLU allow us to appreciate the meaning and sensitivity of utterance length measures in general. Despite Klee and Fitzgerald's (1985) criticism, a large amount of research suggests that there is evidence that MLU is highly correlated with age until about 48 months, as well as with the development of morphological and syntactic skills in English-speaking young children (Blake, Quartaro & Onorati, 1993; Miller & Chapman, 1981; Rollins, Snow & Willett, 1996; Rondal, Ghiotto, Bredard & Bachelet, 1987; Scarborough, Rescorla, Tager-Flusberg, Fowler & Sudhalter, 1991). A strong relationship of utterance length with morphosyntactic skills has been established, at least between MLU scores of 1.00 and approximately 4.50, which is typically reached at about 48 months. Afterwards, utterance length can still provide a general guideline in language development, but it is less clearly related to syntactic development.

Related questions concern the reliability of utterance length measures. Following Brown (1973), rules for calculating utterance length include criteria for sample length (100 utterances, but 50 utterances have been considered as an acceptable sample size). However, the reliability of utterance length was strongly questioned by Klee and Fitzgerald (1985), who addressed the question of intra-sample variability of MLU. From the same perspective, Rondal and colleagues (Rondal *et al.*, 1987) concluded that only the period with average MLU smaller than 3.5 shows acceptable variability. However, in this article, we propose a different view, namely that variation in variability is informative on the nature of developmental change (see also van Dijk, de Goede, Ruhland & van Geert, 2003; van Geert & van Dijk, 2002).

What emerges from all these studies is that utterance length measure is a valid and reliable index of grammatical development for the age range concerned in our study, although it has also been shown to be sensitive to pragmatic influences, such as differences in situation and discourse context (Bornstein, Painter & Park, 2002; Jonhston, 2001). However, the implicit assumption behind these works on MLU is that the relationship between utterance length and grammatical development is continuous. That is, an increase in length corresponds with an increase in syntactic development, up to a point in development where the connection dissolves and utterance length is no longer a reliable indicator of the level of syntactic development. The authors of this article take a different stance, which is more closely related to the discontinuity assumptions of developmental stage approaches. For instance, it is not unlikely that a predominance of one-word utterances in a child's early speech is an indicator of a mechanism of language production that is qualitatively different from the mechanism that corresponds with a predominance of longer utterances. From this perspective, such qualitative changes correspond with discontinuities that are indicators of major transitions in early language and grammatical development. It is important to emphasize that, in our view, the relationship between the assumed underlying production mechanisms on the one hand and utterance length on the other is *probabilistic* and *conditional*. This issue is of considerable methodological importance and will be more fully discussed later in this article.

Few researchers will doubt that in order to answer questions about eventual discontinuities or transitions in early language development, qualitative linguistic analyses of the utterances at issue are indispensable. However, the central aim of this article is to investigate whether a *quantitative* analysis of the data might also contribute to answering this issue.

Theoretical expectations about (dis)continuity in the acquisition of grammar

The questions are, first, how does children's progression in utterance length reflect their progression in grammatical development and, second, is there any evidence that grammatical development takes the form of discontinuous change?

Although there are individual differences in the transition from simple words to multi-word combinations, a process that occurs less clearly in so-called 'expressive style children', the passage from one-word utterances to first-word combinations to multi-word utterances is usually considered to be a general pattern of child language development. At around 1 year of age, children usually begin to produce isolated non-inflected words that may be thought of as 'holophrases' in which a single linguistic symbol is used to convey a relatively undifferentiated communicative intention (Barrett, 1982). A major step in children's developing linguistic competence is the appearance of the first productive word combinations that begin at around 18 months of age: different words are used to indicate different components of a scene (event and object, such as 'More juice'). Thereafter, they progressively produce longer utterances containing linguistic symbols and syntactic structures, such as transitive constructions and complex structures.

A central question raised by this description is when and how children have abstract categories denoting an adult-like syntactic level of language. Many researchers in the 'Universal Grammar' approach assume that young children have the same underlying grammatical competence as adults from the very beginning. In the perspective of this 'continuity assumption', the discrepancy between children's hypothesized linguistic competence and their performance should be attributed to external factors, i.e. to development in domains other than grammatical competence, such as limitations in memory and processing capacities, or to genetic maturation in peripheral aspects of linguistic competence.

In contrast, constructivist approaches to language acquisition (e.g. Bates & Goodman, 1999; Berman, 1986; Lieven, 1997; MacWhinney, 1999; Tomasello, 2000, 2003; Tomasello & Brooks, 1999) assume that abstract linguistic categories are gradually constructed by the child, who goes from immature to adult-like levels of language on a step-by-step basis. From this perspective, Tomasello (2000) emphasized that young children's productivity with language has been grossly overestimated and that 'young children's earliest linguistic productions revolve around concrete items and structures; there is virtually no evidence of abstract syntactic categories and schemas' (Tomasello, 2000, p. 215). For example, studies questioning the acquisition of verbs and verb-argument constructions in English-speaking children's spontaneous speech during their second and third years of life indicate that children generally used some verbs in certain kinds of syntactic constructions, and other verbs in other kinds (e.g. 'the verb island hypothesis'; Tomasello, 1992, 2000; the 'lexically based learning hypothesis'; Lieven, Pine & Baldwin, 1997). A similar view also emerged from naturalistic studies on languages other than English, which emphasized that morphological and syntactic categories develop gradually (Armon-Lotem & Berman, 2003; Bassano, 2000; Bassano, Laaha, Maillochon & Dressler, 2004; Pizutto & Caselli, 1994; van Geert & van Dijk, 2003).

The conception of a gradual process in children's construction of grammars proposes a notion of continuity quite different from that of Universal Grammar. It relies on the idea that grammatical categories emerge and develop from interactions between children's endogenous predispositions and stimulation from the input, as well as from interactions with other language skills, such as lexical capacities. In our view, based on dynamic systems theory, this conception is compatible with the idea that this process shows discontinuities reflecting critical points in the course of development, in the sense that gradual construction takes the form of discontinuous steps.

It is important to note that the notions of continuity and discontinuity bear different kinds of meanings. Two such meanings are primarily of a linguistic and developmental nature. One is that assigned by the framework of Universal Grammar, another refers to the constructivist framework and to developmental (dis)continuities in general. The third is a mathematical meaning that will be used in the remainder of this article and which is associated with particular statistical tests (see van Dijk & van Geert, 2007, for a definition of continuity and discontinuity). It is assumed that evidence of a discontinuity in the mathematical sense displayed in the data lends support to the hypothesis that there is an underlying process that is discontinuous in the developmental sense.

Three steps in early grammatical development?

In the line of research presented above, we hypothesize that hierarchies in the production of utterance types of different length reflect different levels in grammatical development. We propose a grouping that differentiates between three levels in utterance length: one-word (W1), two- and three-word (W2–3) and four-and-more-word (W4+) utterances. We postulate that a strong dominance of W1 utterances is associated with an overall holophrastic (non-combinatorial) stage of language, and a dominance of W2–3 utterances with an overall simple combinatorial stage of language in which abstract syntactic categories are not necessarily required (although, of course, precursors, emergent and even true grammatical devices can be found in these periods). Finally, a dominance of W4+ utterances is assumed to correspond to a more sophisticated stage of grammar in which productive categories (such as noun and verb grammatical classes) and syntactic relational devices (such as auxiliaries, prepositions and conjunctions) are used.

This view can fit the descriptions of one-word utterances and first word combinations in French language in particular (Bassano, Eme & Champaud, 2005; Veneziano, Sinclair & Berthoud, 1990; Parisse & Le Normand, 2000). In general, early W1 utterances are likely to be formed of expressive and socio-pragmatic devices, such as interjections and yes–no particles, little functional words (such as adverbials *voilà*, 'here (is)'; *là*, 'there'; *encore*, 'more'), and certain content words, which are mostly nouns. The early W2–3 utterances may be combinations of content words (such as *papa parti*, 'daddy gone'), or combinations of content words with expressive or socio-pragmatic devices (*non veux pas*, 'no don't want'), or combinations of content words with function words such as determiners or preverbal pronouns (or fillers), which are likely to be formulaic expressions (*ça y est*, 'all done') or simple presentative expressions using the copula *est*, 'is' (*est lela chat*, 'is the/filler cat'; *c'est papa*, 'this is daddy'). Such combinations show the appearance of simple structures that possibly involve emergent morphological devices, but they do not show the emergence of a full-fledged grammar (Veneziano *et al.*, 1990).

In our view, a criterion for the emergence of an adult-like grammar is children's use of full SVO (subject-verb-object) transitive constructions (i.e. prototypical two verb argument structures). In French, these constructions can hardly be contained in W3 utterances, due to the combination of various morphosyntactic constraints. In particular, an obligatory grammatical subject (noun phrase or pronoun) is required in the verb phrase, and an obligatory determiner is usually required in the noun phrase. Moreover, transitive constructions require a verb in a finite form, and, if we put the present form apart, a number of the most frequent finite forms are compound structures (e.g. compound past, periphrastic future), consisting of a grammatical verb and a main verb. For these reasons, full transitive constructions are likely to need more than three words. This point was verified by analyzing W2, W3 and W4 utterances in Pauline's corpus, one of the two children under study. First, we found that the proportion of utterances involving full finite verb forms did not reach 50% for W2 (89/442: 18%) and W3 utterances (155/319: 48%), although it was clearly higher in W3 than in W2, whereas it reached 80% for W4 utterances (153/194). Second, and more importantly, the child's verbal W2 utterances did not contain any full transitive constructions, and only two were found in all her verbal W3 utterances, whereas a number of 41 full transitive constructions were found in the child's verbal W4 utterances.

Relationship between the grouping based on utterance length and underlying generators

The relationship between utterance length and underlying processes of language production is complicated and needs to be delineated as clearly as possible before utterance length can be used in a study of underlying developmental processes. Let us begin with a null hypothesis model that assumes only one underlying generator, S* ('S-star'), which changes across developmental time in a continuous fashion and does not undergo major transformation (the present authors take no standpoint as to which form the underlying S* grammar should take). Under the S* assumption, the growth of average utterance length is, in all likelihood, continuous, with longer utterances occurring more frequently as the child grows older. It can be shown that if average sentence length increases continuously and even linearly, the *proportion* of W1, W2–3 and W4+ utterances will show a pattern that is virtually identical to the pattern predicted by the three-step model discussed earlier, namely a negative S-shape for W1, an inverted U-shape for W2–3 and a positive S-shape for W4+ (the mathematical justification for this claim exceeds the scope of the present article). Thus, the difference in shape of the developmental curves of W1, W2–3 and W4+ utterances does not in itself provide a justification for the assumption that these groups of utterances are produced by qualitatively different generators. However, the above-mentioned three-step hypothesis predicts two transitions accompanied by two statistical anomalies, whereas no such anomalies are predicted by the continuous S* model. Thus, the justification for discontinuity must be found in the finding of the predicted transient anomalies.

In contrast with the null hypothesis of a single generator changing in a continuous fashion, we have thus formulated an hypothesis of three successive generators, the holophrastic, simple combinatorial and syntactic generator respectively, which will show a developmental process characterized by two discontinuities, namely the H-C and the C-S transition. The relationship between the occurrence of particular utterances and the hypothesized generators is hierarchical. That is, the holophrastic generator produces mainly W1 utterances, the combinatorial generator, once present, typically produces W2–3 utterances, but can also increasingly account for eventual W1 utterances. The syntactic generator is needed for W4+ utterances, but can also increasingly account for W2–3 and W1 utterances. Thus, it follows that it is *not* the mere *presence* of, say, a W3 utterance that should be taken as an indicator of the hypothetical C-generator. Rather, it is the *emergence* of W2–3 that will indicate the emergence of the underlying C-generator and the *emergence* of W4+ that indicates the emergence of the syntactic generator.

The hypothetical H-, C- and S-generators emerge and consequently vanish (the H- and C-generators) in the indicated order. It is likely that this emergence and disappearance occurs in the form of overlapping waves (see Figure 1; see Siegler, 1996, for a general account of overlapping waves in development). From this (idealized)

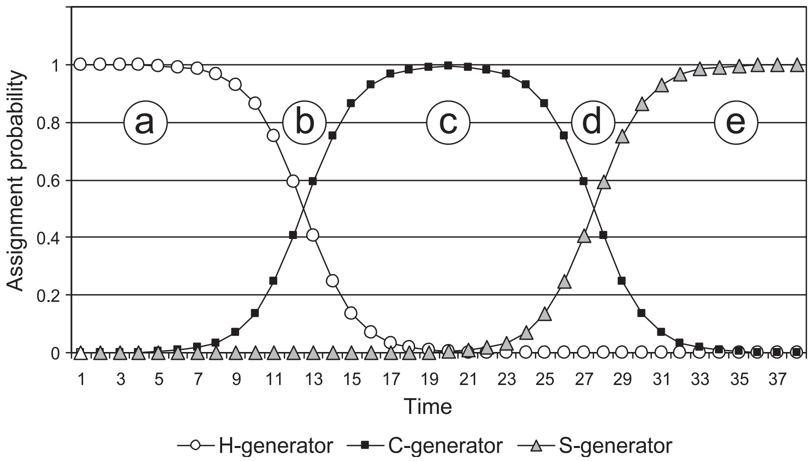


Figure 1 *The holophrastic, combinatorial and syntactic generators, represented as probability waves. Point* a *implies that almost only W1 utterances will be found and that there is a probability of nearly 1 that such utterances can be assigned to the H-generator. Point* b *implies that W1 as well as W2–3 utterances occur at this time and that the probability that a W1 utterance is produced by the H-generator is 0.5 (and 0.5 that it is produced by the C-generator). At point* c*, all utterances are W1, W2 or W3 and are produced by the C-generator only. At point* d*, utterances range from W1 to W4+ length. All W4+ utterances are generated by S, and W1 to W2–3 utterances are generated by the S- or by the C-generator, both with a 0.5 probability. At point* e*, all utterances are generated by S.*

representation, it follows that there are two unstable points (point *b* and point *d* in Figure 1), at which two generators (H versus C and C versus S, respectively) are equally likely to have produced the observed utterance at issue and thus stand in a sort of competitive relationship. Around these time points we expect to find statistical anomalies, such as increased fluctuation or eventually temporary regressions.

In the next section we will discuss a general framework for the study of developmental change, dynamic systems theory, and show how it can help us understand how a simple quantitative criterion such as utterance length can be used to study developmental transitions.

Dynamic systems models: developmental change and variability

Dynamic systems models of developmental processes: definition and properties

The mathematical definition of a dynamic system is 'a means of describing how one state develops into another state over the course of time' (Weisstein, 1999, p. 501). In this particular case, the state of the system is expressed by a single descriptor, utterance length, which refers to the number of utterances of length 1, 2, 3, etc. during a standard observation session. Each such type of utterance (1 word, 2 word, etc.) is a variable in the system. In accordance with the preceding sections, we will collapse the utterance types into three groups, namely the number of 1-word, 2- or 3-word and 4-or-more-word utterances, which, as explained earlier, are potential indicators of hypothesized NC-, C- and S-generators.

The system consisting of the three different types of utterances constitutes a dynamic system, which consists of components that exert specific influences or forces upon one another and by doing so, change each others' and their own properties (van Geert, 2001; for overviews of dynamic systems thinking in development see Thelen & Smith, 1998; van Geert, 2003). Thus, in this particular case, it is assumed that the hypothesized underlying generators affect each other, in such a way that their properties – more precisely the frequencies of the generated utterance types – change over time.

In dynamic growth models (van Geert, 1994; Fischer & Bidell, 2006), the relationships between the major order variables, corresponding with the utterance types and associated generators, can be either neutral (no relationship), supportive, competitive or conditional. A supportive relationship means that the level of the supporting variable has a proportional and positive effect on the growth of the supported variable. A competitive relationship means that the level of the competitor has a proportional and negative effect on the growth of a variable, i.e. that it contributes to the latter's decline. Finally, a conditional relationship implies that a certain level of a conditional variable is needed in order to get the growth of a dependent variable off the ground.

Indicators of critical points and transitions

A dynamic model based on these relationships will produce a pattern of continuous changes in the variables involved, simply because the mathematics in which those models are formulated is a mathematics of continuous functions (this is not a logical requirement or limitation, however; see van der Maas & Molenaar, 1992). However, the actual system may pass through critical points or stages that characterize major changes in the underlying dynamics. Such qualitative changes or transitions fall outside the scope of the currently discussed growth dynamics and require a different modeling approach. The critical points are often characterized by significant changes in the form of the fluctuations that are characteristic of the system and can be discovered by analyzing changes in the variability pattern over time.

According to catastrophe theory (see van der Maas & Molenaar, 1992, for an application to developmental data; Ruhland & van Geert, 1998), critical points and transitions are characterized by discontinuities, which leave a number of observable marks on the data, the so-called catastrophe flags. One of these flags is anomalous variance, which refers to the fact that the pattern of variability in the vicinity of a discontinuity is different from that before and after the developmental change. Van Dijk and van Geert (2007) have taken a different stance, focusing on the notion of transition rather than discontinuity *per se*. Transitions may occur in the form of continuous patterns of change that are accompanied by various sorts of anomalies, such as unexpected but transient peaks in the frequency of a phenomenon, changes in the pattern of fluctuations, i.e. intra-individual variability or temporal regressions (see also van Geert & van Dijk, 2002). The anomalies are conceived of as indicators of a possible underlying discontinuity.

Intra-individual variability and transitions

If a new generator emerges – for instance the combinatorial generator that is supposed to succeed the holophrastic one – the system will for some limited time be in a bimodal attractor state (both generators are possible). Over the short term, it will fluctuate between these two possible attractor states. The variability characteristic of the original state (e.g. the holophrastic generator) will be temporarily mixed with the variability characteristic of the new generator, and this mix will be observable in the form of temporarily increased variability. If we show that a peak in variability is associated with an otherwise continuous passage to the dominance of a certain kind of utterance type (e.g. from W1 to W2–3, or from W2–3 to W4+), we will have provided supporting evidence for the claim that the corresponding period is a critical point in the development of grammar. Since we assume that W1, W2–3 and W4+ utterances probabilistically refer to three different underlying generators, we expect two points of increasing variability. The first will coincide with the transition from a dominant W1 to W2–3 mode, the second with the transition from a W2–3 to a dominant W4+ mode.

Data

Participants

The main set of data used in this article came from the longitudinal corpus of one French girl, Pauline, who was the youngest of four children in a middle-class family living in Rouen, a mid-size city, and who was studied from 1;2 to 3;0. In addition, we also used data coming from another French child, a boy named Benjamin, who was the younger of two children in a middle-class family living in Paris, and who was studied from age 2;0 to 3;0. Previous studies conducted on these children (e.g. Bassano, 1996, 2000; Bassano, Maillochon & Eme, 1998; Bassano *et al*., 2004) showed that they both had normal language development, with evidence of productive language by the middle of their second year.

Data collection[1] and data sampling

Data were obtained using a free speech sampling method, in which the child's naturalistic productions were systematically collected and transcribed. Each child was audio- or video-recorded at home, during everyday activities, such as eating, playing, washing, dressing, etc., in unstructured interactive sessions with her/his family. Pauline was recorded twice a month, during sessions of about 2 hours each. Long uninterrupted parts of each recorded session were selected for transcription so as to obtain a variety of situations and a sufficient and representative number of productions. Data collection was less regular for Benjamin, who was sometimes recorded once a month and sometimes three or more times a month, during sessions varying in length (from 10 minutes to 2 hours). For both children, transcriptions were made in accordance with CHAT format (MacWhinney, 2000). The child's verbal productions were fully transcribed (orthographically and eventually phonetically), as well as the other participants'

[1] Data collection took place between 1991 and 1994 for Pauline (collection and transcription by I. Maillochon, control by D. Bassano). For Benjamin, it took place between 1987 and 1989 (collection and transcription by D. Bassano and J. Weissenborn and collaborators).

productions, and indications were provided about the situations, contexts and gestures.

Following the basic principle generally used in our longitudinal naturalistic studies (Bassano, 2000; Bassano *et al.*, 1998, 2004), analyses were first conducted using chronological months as units, eventually combining two sessions (for Pauline's data). Monthly samples consisted of a constant number of 120 utterances. Samples were all the more naturally and representatively constituted, since they were formed by all the utterances of several long and uninterrupted discursive sequences. As far as possible, the discursive sequences were chosen balancing the contexts and situations (e.g. playing, eating, etc.), and excluding songs and imitation games. As is customary, utterances were defined as the vocal productions that were prosodic and meaningful units, including at least one element resembling a French word in form and meaning. Babbling, vocalizations and completely incomprehensible productions (but not repetitions) were excluded from the samples.

In order to analyze intra-individual variability, the monthly 120 utterances samples were divided into two equal sub-samples, one formed of the first 60 utterances and the other one formed of the following 60 utterances. For Pauline's data, and for the period from 14 to 29 months, each of the two sub-samples corresponded to each of the two distinct sessions recorded during the month; for the period from 30 to 36 months, the two sub-samples came from the same session (in fact, only one of the two monthly sessions was transcribed for this period). For Benjamin's data, the two samples came from the same session. Although we are aware that the origin of the two sub-samples of 60 utterances from distinct sessions in Pauline's data for the period from 14 to 29 months was somewhat problematic for the analysis of variability, we could not change these data which had been collected and submitted to coding as they were in a larger research program. Solutions to this problem are proposed in the section that analyses the variability patterns. Finally, the 60-utterance blocks were further subdivided into 30-utterance blocks in order to enable us to analyze the within-session variability.

Utterance length measures

Utterance length was assessed by coding the number of words calculated for each utterance. Following recent recommendations (Hickey, 1991; MacWhinney, 2000; Thordardottir & Ellis Weismer, 1998), we used words (that is, free morphemes, and not morphemes), because this count seems the most simple and appropriate measure. Moreover, a strong correlation (around .98) was found between MLU in words and MLU in morphemes in a number of studies. Using a count in words means that verb inflections were not counted separately (*mange*, 'eat', was counted as one word, *veux manger*, 'want to eat', was counted as two words). By contrast, but following the same logic, that is using the more simple and standard notion of word count, amalgams and formulaic expressions (e.g. *au revoir*, 'bye bye'; *ça y est*, 'all done'; *s'il te plait*, 'please') were counted for as many words as there were free morphemes (distinctly produced by the child) in the lexical unit, although these expressions were considered as specific units in other types of coding, such as lexical analyses. A result of these decisions in coding utterance length was that the number of words per utterance was generally maximized.

Utterance length coding was made according to 'raw' and 'net' versions, the latter removing incomprehensible and tentative words. We used the net version of the word count, because this index is more appropriate for assessing a progression in grammatical development. In using the net version of the word count, we can assume that increase in utterance length and increase in variability were not due to a limitation in speech performance and/or an increase in speech errors. We excluded from the net word count incomprehensible elements which could be produced in the course of an utterance (e.g. *xxx carottes*, 'xxx carrots', was counted as one word), as well as false starts, that is words repeated because of the child's hesitations in forming utterances (e.g. *il est, il est bleu*, 'it is, it is blue', was counted as three words). There is another kind of element that can provide a decision problem in word counting: the so-called 'filler-syllables' (also called Prefixed Additional Elements, see for instance, Veneziano & Sinclair, 2000). Although the phonological versus syntactic nature of the knowledge implied by the use of fillers is discussed (Peters, 2001), we included fillers in the 'net' word count because these elements were generally considered as placeholders for grammatical morphemes.

On the bases mentioned above, each child's MLU, as well as the frequencies of one-word, two-word, three-word utterances, etc. (W1, W2, W3, etc.) were calculated for each monthly sample (120 utterances) and sub-sample (60 utterances and 30 utterances). The longest utterance in Pauline's data was a W17, and a W21 in Benjamin's data (at 36 months for each child).

Analysis I: Developmental changes in the process of increasing utterance length

The development of W1, W2–3 and W4+ utterances

We first examine how each of the three utterance type groups develops over time in the two children's data. Figures 2a and 2b show the development of W1, W2–3 and W4+ utterances in Pauline's data from 14 to 36 months.

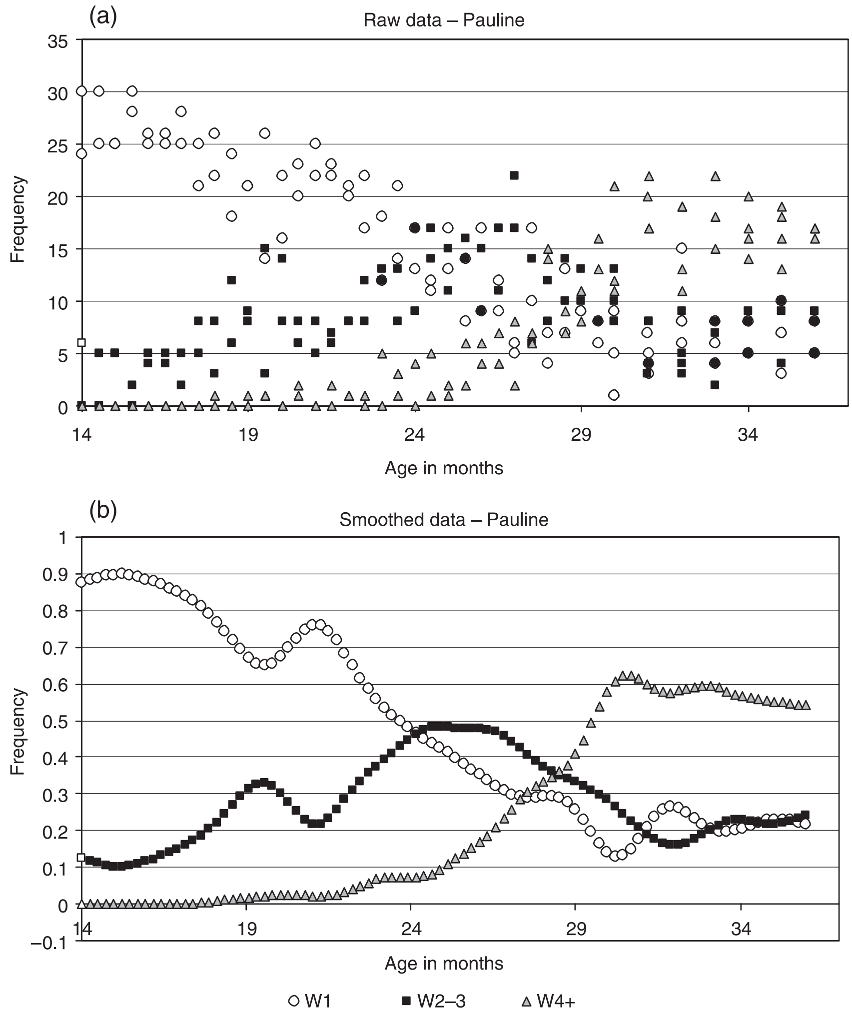
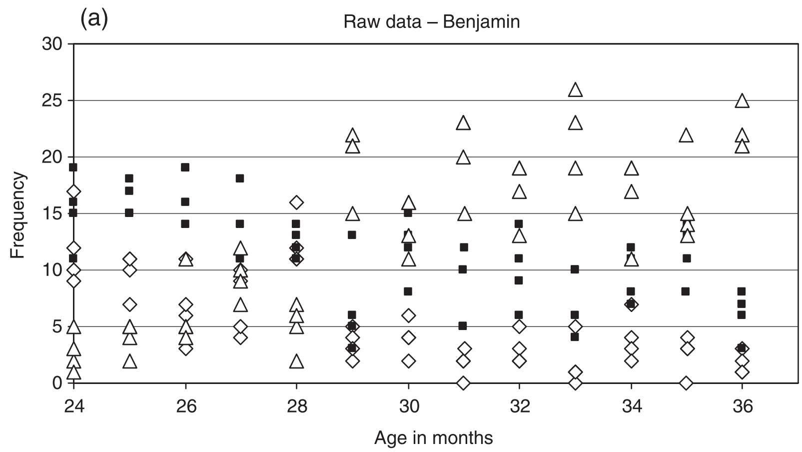


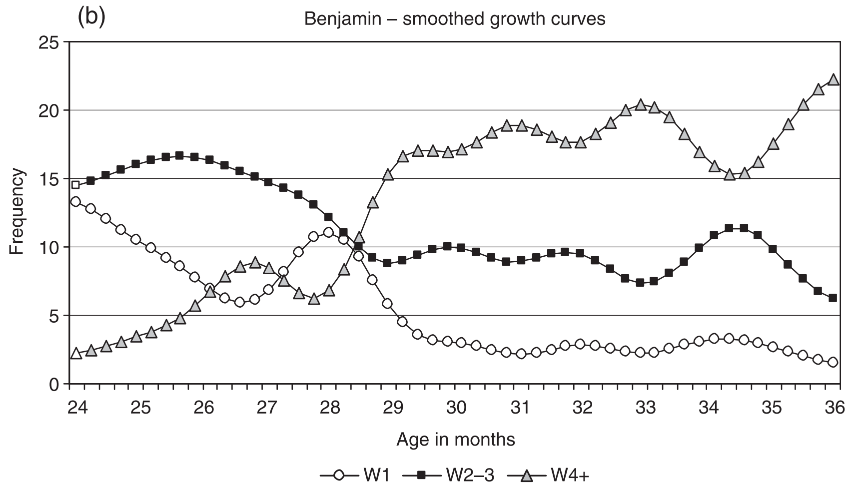
Figure 2 *W1, W2–3 and W4+ utterances of Pauline: raw data and smoothed curves.*

Figure 2a shows the raw data, i.e. the number of utterances in each group for each sub-sample of 30 utterances (a complete observation consisted of 60 utterances and was split in half in order to be able to calculate within-session variability). Figure 2b shows the smoothed curves of the raw data, based on the Loess smoothing procedure. Loess smoothing, i.e. locally weighted least-squares smoothing, applies a moving window across the data, and fits a regression model over the window by weighting the data proportional to their distance from the middle of the window (Siminoff, 1996; Härdle, 1991; the smoothing was done in Table Curve 2D). The window was set to an equivalent of 3 months, which gave the best compromise between a smoothed curve and symmetric residuals. As shown in the figure, Pauline's W1 utterances were at the top level at the beginning of the observation (14 months) and showed a declining trajectory until they leveled off after the 30th month. The W2–3 utterances presented an inverted-U-curve showing a temporal regression at around the 21st month and a peak at around the 26th month (half of the utterances during a session), and began to stabilize after the 32nd month. From the W2–3 peak on, the W1 and the W2–3 frequencies were very close to one another. The W4+ utterances showed an S-shaped increase: they started off around the 23rd month, began to dominate upon the W1 and W2–3 utterances around the 28th month and reached a maximum level at around the 31st month (more than half of the utterances during a session).

Figures 3a and 3b show the development of the three utterance-type groups in Benjamin from 24 to 36 months, Figure 3a the raw data, and Figure 3b the smoothed data.

Since the data collection started later for Benjamin than for Pauline, Benjamin's data capture only the more advanced part of the developmental trajectories. At the start of the observations (24 months), Benjamin's W1





Figures 3 *W1, W2–3 and W4+ utterances of Benjamin: raw data and smoothed curves.*

utterances were already in a later stage of decline, but they leveled off at about the same time as those of Pauline, although their final level was a little lower. The W2–3 utterances had almost reached their top in the 24-month observations (about half of the utterances in a session), started to decline around the 26th month as for Pauline, temporarily stabilized after the 29th month (about one-third of the utterances in a session), showing a final decline in the last observations. At the beginning of the observations, the W4+ utterances were in their increasing trajectory, started to dominate over the W1 and W2–3 utterances around the 28th month, reached a maximum at around the 32nd month (about two-thirds of the utterances) and finally showed a new increase in the last observations.

In summary, the analysis of how the W1, W2–3 and W4+ utterances evolve shows the structural similarity of Pauline's and Benjamin's linguistic production. Although Benjamin's data captured only the last part of the trajectories, all three developmental patterns, as well as their relations, are strikingly similar in the two children. In particular, the W4+ utterances showed a perfectly similar increase, starting to dominate both the W1 utterances and the W2–3 utterances from about 28 months on in both children and increasing rapidly thereafter.

A dynamic growth model of W1, W2–3 and W4+ utterances

The growth model

The dynamic growth model conceives of W1, W2–3 and W4+ utterances as distinct attractor states of the short-term dynamics of utterance production to the extent that they are based on distinct underlying processes and mechanisms, i.e. 'generators' (see Introduction). A generator refers to the child's 'linguistic knowledge', procedures or ability needed for the actual production of W1, W2–3 and W4+ utterances and to all aspects of the environment that help the child produce such utterances.

Although we do not know the exact properties of these generators, we believe it is nevertheless possible to postulate general, long-term relationships among them,[2] which are either supportive, competitive or conditional. They apply to the growth or decline of each of the generators, as probabilistically expressed by the growth or decline of the production of the corresponding W1, W2–3 and W4+ utterances.

The dynamic model assumes, first, that 'simpler' forms have a supportive relationship with the more complex forms. Thus, the holophrastic W1 will have a supportive relationship with the combinatorial W2–3, and W2–3 will support the syntactic W4+. Second, we assume that there exists a conditional relationship between the simpler forms and the more complex successors. The conditional relationship takes the form of a critical mass. In principle, the relationship will hold between W1 and W2–3 and W2–3 and W4+, respectively, but since W1 has a maximal value at the beginning of the series of observations, it makes no sense to simulate the corresponding conditional relationship. Third, we assume a competitive relationship between the more complex and the simpler form, i.e. a competitive relationship from W2–3 to W1 and from W4+ to W2–3, respectively. The assumption of a supportive relation (W1 with W2–3 and W2–3 with W4+) is backed up by various arguments. Simpler forms eventually supply the more complex forms with constituents (of various kinds) and probably have a positive effect on the child's sensitivity for linguistic forms of higher complexity as used by other speakers in the environment, etc. The conditional relationship (both between W1 and W2–3 and between W2–3 and W4+) can be explained by the argument that a critical mass of simpler utterances is required for the emergence of more complex utterances: the child must be able to produce a certain amount of simple combinations of different kinds before being able to produce complex syntactic constructions and utterances. This explanation is consistent with the 'critical lexical mass' hypothesis, which claims that development within morphosyntax is triggered by an increase in the size of the lexicon beyond a given level (e.g. Marchman & Bates, 1994; Bates & Goodman, 1999; Bassano, 2000). Finally, a possible justification for the competitive relation (from W2–3 to W1 and from W4+ to W2–3, respectively) is that the more complex form is more strongly supported by the environmental input (it resembles the language of the linguistically mature speakers) than the simpler form and thus automatically leads to the decline of the latter.

[2] They are 'long-term' in contrast to the short-term dynamics of utterance production, and in principle comprise the time window of our study, for instance the 22 months during which Pauline's utterance production has been recorded.

The aim of the simulation is to obtain the best possible fit of the smoothed data curves, given the dynamic relationships among the variables as described in the previous section. The smoothed data curves are used because they represent the developmental changes in the *likelihood* or probability that the child will produce a W1, W2–3 or W4+ sentence. The current growth model is not aimed at – and also is not suited to – explaining the short-term dynamics of utterance production, which, for instance, would entail simulations of the variability in the actual numbers of produced utterances. The dynamic model is based on the logistic growth model with added supportive, competitive and conditional relationships (see van Geert, 1991, 1994). It has been implemented in the form of a program written by the second author and is running under Microsoft Excel. The estimation of parameter values that best fit the data is based on an optimization procedure, which also runs under Excel (which is part of the Poptools Excel add-in; Hood, 2004).

Simulation results

The dynamic model has been fitted to the smoothed data of Pauline and Benjamin, which have been transformed into 300 evenly spaced (interpolated) data points (Figures 4 and 5).

The parameter values are shown in Table 1 for Pauline's model. Since Pauline's data cover virtually the whole range of changes in the three types of utterances, the estimated parameters are more valid than those of Benjamin (in Benjamin's case, the optimization procedure operates on a subset of the required data and finds optimal values within those limits, which do not necessarily correspond with the values that would have been found if the data had started at a significantly earlier age). The analysis will be confined to Pauline's model and start with the fit of the observed frequencies (uncorrected model).

The W4+ utterances have a considerably greater growth rate than the others, which suggests that the growth of W4+ utterances looks more like an actual transition process than the growth of the others. Characteristic of a transition is a period of virtually no change followed by rapid increase. The support/competition parameters confirm the model explained earlier. The simulation supports the three assumptions: (1) that simpler forms have a supportive relationship with the more complex forms (W1 with W2–3 and W2–3 with W4+); (2) that there exists a conditional relationship between the simpler forms and the more complex successors (as was the case between W2–3 and W4+); and (3) that there exists a competitive relationship between the more complex and the simpler form (from W2–3 to W1 and from W4+ to W2–3, respectively). Finally, the

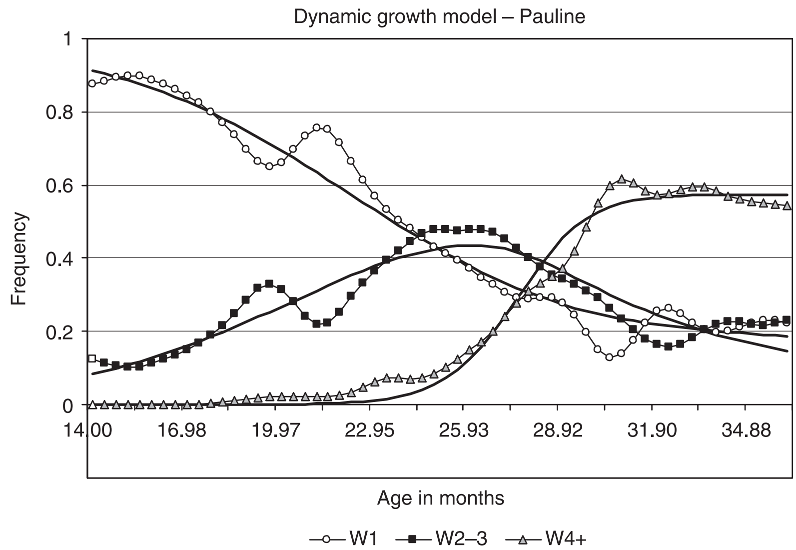


Figure 4 *A dynamic growth model of Pauline's data.*

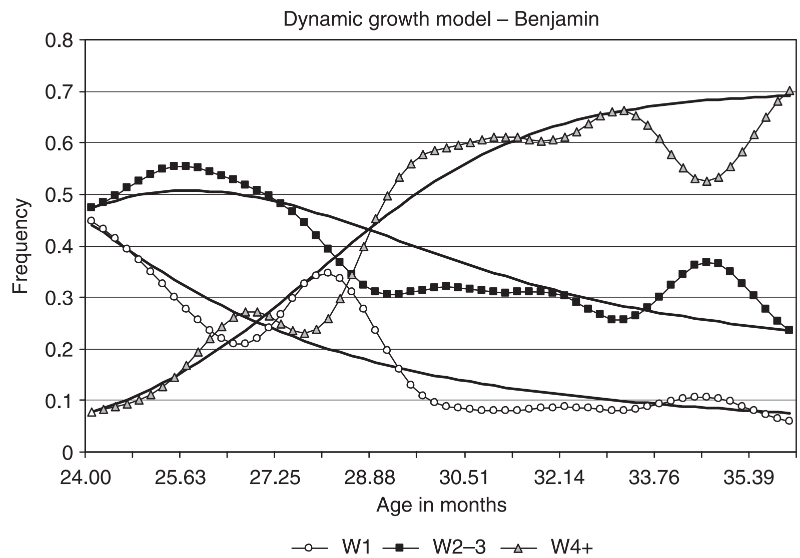


Figure 5 *A dynamic growth model of Benjamin's data.*

R^2 values show that the fit between model and data is very good.[3]

It might be objected that the dynamic model does not describe the frequencies *per se*, but that it models the productivity of the generators that are assumed to be responsible for the production of particular utterance types. Since the syntactic generator can produce sentences of any complexity, it is likely that as it becomes more fully established, it will also produce an increasing proportion of the W1 and W2–3 utterances. In line with this assumption and with a procedure defended elsewhere (van Dijk & van Geert, 2005), we propose a scenario in which the utterances produced by the hypothesized holophrastic and combinatorial generators (the W1 and W2–3 utterances) continue to decline after month 31, instead of stabilizing as the actual numbers of W1 and W2–3 utterances do. The qualitative pattern of parameter values of this corrected model is similar to that of the uncorrected model, in particular as far as the relationships between the utterance types are concerned.

[3] Note that the estimated parameter values for Benjamin's data were qualitatively similar to those of Pauline. As expected, the fit of the model was a little worse than that of Pauline's (R^2 of 0.8, 0.7 and 0.92, respectively).

Table 1 *Estimated parameter values for the dynamic growth model based on the original data of Pauline and on Pauline's corrected data (which are values between brackets)*

Growth	W1	W2–3	W4+
Growth rate	0.03 (–0.01)	0.09 (0.11)	0.78 (0.37)
Initial value	0.91 (0.92)	0.09 (0.06)	0.01 (0.03)
Carrying capacity	1 (1)	0.35 (0.65)	0.57 (1)
Support/competition	to W1	to W2–3	to W4+
from W1		0.18 (0.21)	
from W2–3	–0.31 (–0.21)		0.04 (0.08)
from W4+	0	–0.36 (0.65)	
Conditional	to W1	to W2–3	to W4+
from W2–3			0.34 (0.29)
R^2 fit	0.97 (0.97)	0.83 (0.93)	0.99 (1)

Note: *Growth rates* smaller than 0 represent decline, growth rates greater than 0 represent increase; high growth rates (e.g. greater than 0.5) correspond with rapid increase. Estimated *initial values* are close to the real initial values (which they should be if the model fits well). *Carrying capacities* approaching 1 correspond with the theoretical possibility that all utterances (at some point in time) are produced by the corresponding generator if no additional influences are present (e.g. the holophrastic generator in the case of W1 utterances, if support or competition from W2–3 or W4+ is absent). Negative *support/competition* parameter values correspond with competition, positive values with support. The *conditional* parameter refers to the minimum proportion of W2–3 utterances needed for the onset of W4+ utterances. All estimated parameter values are consistent with the theoretical predictions (see text).

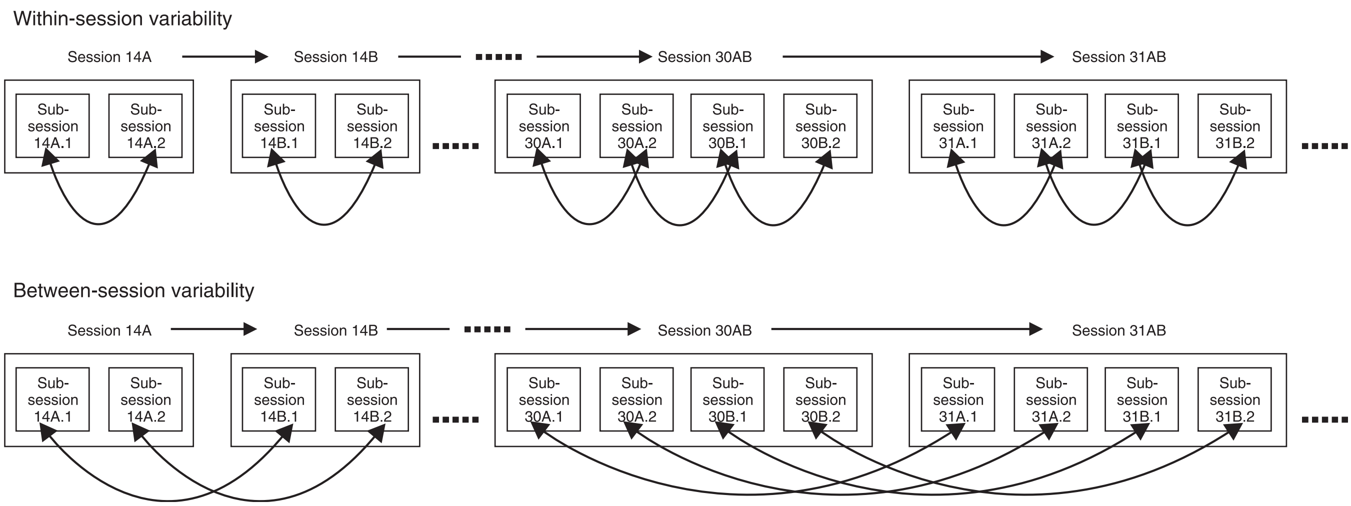


Figure 6 *An illustration of the methods for counting within- and between-session variability for two types of observations, one containing two, the other containing four subsessions.*

In summary, this section shows that the growth of W1, W2–3 and W4+ utterances in Pauline and in Benjamin can be simulated by means of an asymmetric support–competition dynamics, characteristic of the development of qualitatively more complex forms out of qualitative simpler ones.

Analysis II: Intra-individual variability and the occurrence of critical points and transitions

Forms of variability: residual variability and within- and between-session variability

In this analysis, we will focus on three types of intra-individual variability. *Residual variability* is defined as the distance between an expected value and the observed value (e.g. the expected versus observed frequency of W1 at time t). The expected values are specified by the curves based on the Loess fitting procedures described in Analysis I. *Within- and between-session variability* refers to the distance (absolute difference) between an observed frequency and the preceding value of that frequency. Within-session variability is the difference between consecutive subsets of 30 utterances (e.g. its number of W1s) collected during one observation session.

For Pauline's data (see Figure 6 for an explanation of the principle of within- and between-session comparisons) the number of within-session comparisons from months 14 to 29 is 1 (the second minus the first subset) and from months 30 to 36 the number of within-session comparisons is 3 (the fourth minus third subset, third minus second and second minus first). This results in a total of 53 within-session comparisons for each utterance type. Between-session variability is the difference between an observation and the preceding observation, for instance between the number of W1s at month 16 and W1s at month 15.5. With Pauline, there are two subsets of 30 utterances for the observations from 14 to 29 months and four subsets of 30 utterances from 30 months on. Between-session comparisons are made for corresponding subsets, for instance 16a is compared with 15.5a and 16b is compared with 15.5b. This results in 88 such comparisons for Pauline. This procedure for calculating within- and between-session variability has also been followed with Benjamin's data and results in 39 and 48 comparisons, respectively.

The division in between- and within-session variability is justified as follows. Given the rate of change of early language development, it is possible that between-session variability, especially if the observation interval is as long as 1 month, is highly correlated with the rate of change of the observed variable. That is, between-session variability is likely to confound variability in the sense of fluctuation with variability in the sense of real change. On the other hand, within-session variability potentially poses its own specific limits. For instance, if during an observation the child is in some sort of 'W1-mood', it is likely that it will last for a good part of the total observation, thus biasing the number of observed W1s. If this is so, within-session variability will be strongly auto-correlated and thus may easily underestimate the fluctuations that can eventually be observed over intervals that are just a little longer (e.g. day-to-day variability). If all this is true, we might expect that on average between-session variability is considerably greater than within-session variability, and thus that they represent different sources (and kinds) of variability.

To check for this possibility, the averages, standard deviations and autocorrelations (lag 1) for the between- and within-session variability in Pauline's and Benjamin's data were calculated (Table 2).

Against our expectation, in Pauline's data the within- and between-session variability averages and standard deviations are virtually similar. Within-session variability shows the expected autocorrelation, which is, however, quite low. In Benjamin's data, within- and between-session variability differ, although they are of a similar order of magnitude. In summary, the data do not support the assumption that between- and within-session variability shows major differences. However, since within- and between-session variability represent different time scales (hours versus weeks), they will be treated separately in the statistical analyses. Visual inspection of the smoothed within- and between-session variability data (which will be shown later) suggests peaks in variability roughly coinciding with the temporal dip in W2–3 utterances and rapid growth of W4+ utterances, respectively, as expected on the basis of the transition hypothesis. The main question that will be addressed in the following sections is whether these peaks in variability represent real changes in variability.

Table 2 *Comparison of between- and within-session variability*

Pauline	Average	Standard deviation	Autocorrelation
Between-session	9.45	5.75	0.17
Within-session	9.51	5.39	0.30
Benjamin	**Average**	**Standard deviation**	**Autocorrelation**
Between-session	12.4	8.4	0.5
Within-session	10.2	4.9	0.3

Statistical method

The null hypothesis model

In order to test whether the observed variability fluctuations are due either to underlying change or to chance, we need a reasonable null hypothesis model. The null hypothesis model assumes that the observed frequencies refer to an underlying generator (whatever its nature), which at any time corresponds with a specific probability that an utterance of a particular type will be produced (see Borsboom, Mellenbergh & van Heerden, 2003, 2004; van der Linden & Hambleton, 1997; see Lord & Novick, 1968, and Traub, 1994, relating this to true scores). Under the assumption that we can sample the knowledge or the behavior a great many times, the average occurrence of that knowledge or behavior will approach the underlying probability distribution. The variability of the answers given is the standard error, which is determined by the particular probability and by the number of samples drawn from it.

If this model is applied to the occurrence of W1, W2–3 and W4+ utterances, it is assumed that at any given moment, each of the utterance types has a particular probability of occurrence. The sum of these probabilities must add up to 1. These probabilities are given by the values of the smoothed curves of the W1, W2–3 and W4+ utterances, as described in earlier sections. Since the probabilities specify the chance that an utterance of type W1, W2–3 or W4+ will occur, the statistical distribution of those utterances is specified by the properties of a multinomial distribution with given probabilities and sample size (which in the case of a complete observation session is 60 and in the case of a split session is 30). Hence, under the null hypothesis, the variability of the frequencies of W1, W2–3 and W4+ utterances is defined by the properties of a multinomial distribution, which can be calculated (or simulated). Given this null hypothesis, it can be statistically tested whether the observed fluctuations in variability can be subsumed under this random model or not. More intuitively formulated, the statement 'The fluctuations observed in the data set are just accidental' means 'The fluctuations observed in the data can, with reasonable probability, be explained by a multinomial model of probabilities that follow the long-term development specified by the smoothed frequency curves.'[4]

Statistical simulation of variability

Monte Carlo procedure

The variability under the null hypothesis can be simulated as follows. For simplicity, the explanation will focus on the 39 observation sessions of Pauline, containing 60 utterances each. Let us take Pauline's data on month 26 as example. The values of the smoothed W1, W2–3, and W4+ curves are 22.9, 28.6 and 8.3, respectively, which, with some errors of smoothing, adds up to 60 utterances as expected. The proportions of the utterances are 38%, 48% and 14%, respectively. Under the null hypothesis, the distribution of possible frequencies of W1, W2–3, and W4+ utterances at month 26 is the multinomial distribution for a sample of 60 cases and respective probabilities of .38, .48 and .14. Thus, for month 26 and hence for any point in time, from month 14 to 36, we have an expected frequency and an expected range of statistical variability. That is, for each expected frequency, a random sample can be drawn of values for the W1, W2–3, and W4+ utterances. If many such sets are drawn, the same measures of variability that were calculated for the observed frequencies can be calculated for the simulated frequencies. For instance, for each randomly sampled series of sets of 60 values, the within-session or between-session variability is calculated. By thus simulating variability under the null hypothesis, the 95% level of within- or between-session statistical variability can be calculated. If the observed variability level exceeds the 95% level, we know that the chances that this variability level is generated by the null hypothesis model is equal to or smaller than 5%. However, since (in Pauline's case) the number of simulated cases is at least 39 (if only the actual observation sessions are simulated) and at most 92 (if all subsets of 30 utterances are simulated), we need to calculate the probability that the number of times the observed variability actually crosses the 95% boundary can be caused by chance alone. This probability can be estimated by means of the same Monte Carlo technique that was used to estimate the 95% boundary (see for an introduction to Monte Carlo methods, Manly, 1997).

Statistical indicator of variability

In order to express variability, we need a valid summary measure that is also sensitive to developmental changes in the variability that we expect to occur. Intuitively, the (positive) extremes in the variability seem the most interesting or informative indicators of variability over a certain time window. The extremes can be specified by taking the average of the biggest variability values over a moving window, or by alternative methods such as the moving standard deviation of variability, or by the moving 70th percentile. A major advantage of the Monte Carlo procedure is that it can be used to estimate the statistical properties of any chosen indicator that can be numerically expressed. The summary measure of variability chosen in the current investigation is the average of the three or four largest variability values over a moving window of 9 to 13 observations, the number depending on the number of observations. All statistical simulations were carried out in Microsoft Excel, with a statistical add-in that performs random sampling (Poptools, a set of statistical tools for population studies by Greg Hood, 2004, which contains, among others, reliable procedures for random number generation) and a number of Visual Basic functions for the Monte Carlo procedures written by the second author.

Statistical test of variability – Pauline

Within- and between-session variability – Pauline

Figure 7 (top) shows the change in within-session variability compared with the 95% boundary, which was obtained by randomly drawing values from the multinomial distributions specified by the estimated frequencies at each point of observation. The observed variability values clearly exceed the boundary in two places. They occur simultaneously with the temporal regression in the growth of the W2–3 utterances (and the corresponding temporal increase in the W1 utterances), on the one hand, and with the W4+ utterances rapidly growing and reaching stability on the other hand. In order to estimate the probability that such periods of augmented variability (above the 95% level) are due to chance alone, we calculated how many times the random sampling procedure

[4] Note that under this statistical model, issues relating to the reliability of 30 versus 60 (or more) observed utterances do not affect the testing: the null hypothesis model is tested under the same conditions as those that hold for the observations. Smaller numbers of observations will thus be compensated for by greater confidence intervals.

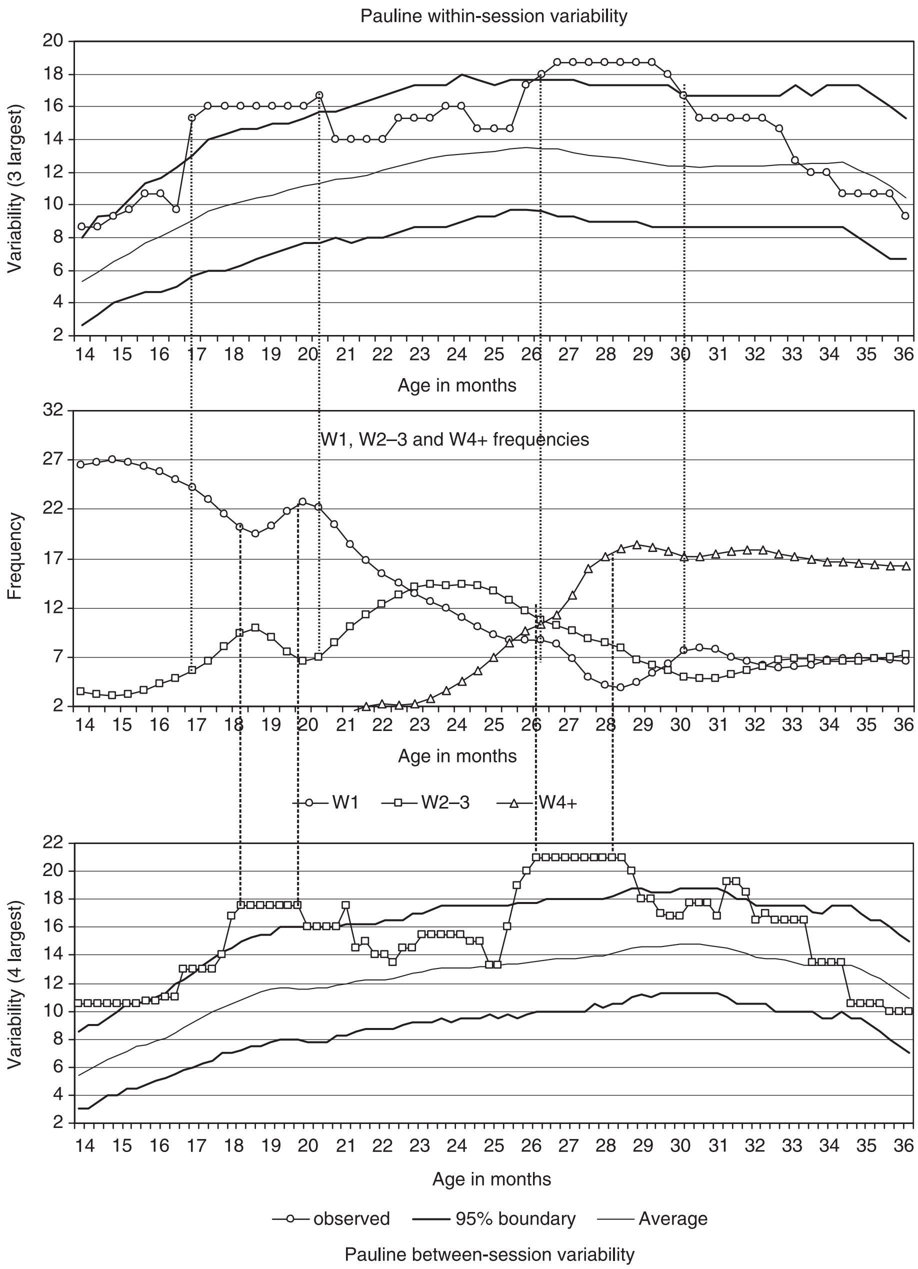


Figure 7 *Total within- and between-session variability of Pauline's W1, W2–3 and W4+ utterances, compared with smoothed frequency curves.*

based on the multinomial probabilities generated two stretches of values that exceed the 95% boundary. Of these two intervals, if any occurred in the simulated set, the length, the sum of values and the average value were calculated and compared with the corresponding observed values. The number of times the simulated sets are equal to or greater than the observed sets (exceeding the 95% boundary) divided by the number of statistical simulations gives an estimation of the *p*-value (see Table 3).

The probability that an interval exceeding the 95% boundary with a length similar to the longest observed interval occurs on the basis of chance (somewhere between months 14 and 36) is 6.3%, that is, the *p*-value of this possibility is .063. However, the probability that a second interval occurs comparable to the second longest observed interval is only 0.2% (that is, a *p*-value of .002). The probability that two intervals such as the observed ones occur on the basis of chance is also 0.2%. In sum, it

Table 3 *P-values for Pauline's variability data*

		Longest interval	Second longest interval	Both longest and second longest interval	Sum of intervals
p-values within-session variability Pauline	length	0.063	0.002	0.002	0.003
	sum of interval	0.047	0.098	*n.a.*	*n.a.*
	average	0.02	0.064	*n.a.*	*n.a.*
p-values between-session variability Pauline	length	0.095	0.013	0.012	$p < .001$
	sum of interval	0.027	0.016	*n.a.*	*n.a.*
	average	0.018	0.007	*n.a.*	*n.a.*
p-values residual variability Pauline	length	0.07	0.04	0.02	0.05
	sum of interval	0.03	0.01	*n.a.*	*n.a.*
	average	0.05	0.13	*n.a.*	*n.a.*

Note: The p-values represent the probabilities that the null hypothesis model produces intervals above the 95% boundary that are as long as or longer than the longest and second longest of such intervals as obtained with the data; additional p-values check the probability of combined intervals ('both longest and second longest interval' and 'sum of intervals'); *n.a.* means 'not applicable'.

is highly unlikely that the observed peaks in within-session variability can be explained by the null hypothesis model.[5]

The pattern of peaks in between-session variability (Figure 7, bottom) is comparable to that of the within-session data (Figure 7, top). Between-session variability shows somewhat more extremes in the high region than within-session variability does. The longer duration of the within-session peaks is probably due to the smaller number of observations (53 versus 88). Given the overall pattern of p-values of between-session variability in Table 3, in addition to the smallest p-value ($p < .001$), it can be concluded that the between-session variability peaks are statistically significant and that they are about similar to the within-session variability peaks. Note that the difference between the simulated and observed between-session variability is not an artifact of the variability in growth rate in the observed data. The simulated variability is based on the growth pattern specified by the smoothed frequency curves and thus automatically reckons with changes in the observed levels.

Residual variability – Pauline

Residual variability is defined as deviation from an expected score, i.e. as the absolute difference between an observed level and the level expected on the basis of the smoothed curve. Since the test of within- and between-session variability supported the predicted peaks of variability, the current test can be viewed as an attempt to check to what extent the findings are dependent on the nature of the variability definition.

As described earlier, residual variability is calculated for the time series of 30-utterance blocks.[6] Figure 8 shows the curves for the estimated 5% and 95% intervals and the observed residual variability measure. As expected, the peaks in the variability are compatible with both the within- and between-session peaks, but are in particular more similar to the between-session peaks.

Table 3 (bottom part) shows the p-values for the distinct criteria, as described earlier. Here, the overall pattern of p-values, in particular the smallest p-value, suggests that the probability that the observed peaks are due to chance alone, given the form and fluctuations in the smoothed frequencies, is below the standard 5% boundary.

How sensitive are the results to the utterance groupings?

Earlier we explained that the relationship between the utterances (W1, W2–3 and W4+) on the one hand and their postulated generators (holophrastic, combinatorial and syntactic) is probabilistic and also changing (see

[5] Note that, in view of the observed peaks of variability, the average observed within-session variability is considerably greater than the simulated within-session variability (about 27% greater); note also that if variability is calculated for the 39 observations only and variability is expressed as the sum of the distance from the expected values, the average observed and average simulated variability are about similar.

[6] Before running the statistical test, it was checked whether the smoothing procedure is sensitive to the format of the data. It was found that the smoothing on the basis of the 30-utterance blocks was qualitatively similar to the smoothing based on the 60-utterance blocks. That is, the form of the frequency curves was virtually identical (that is, if the 30-block smoothing is multiplied by 2, the curves almost completely coincided with those of the 60-utterance blocks). Moreover, since the within- and between-session variability were of the same order of magnitude (see the introduction to the section Analysis II), a different smoothing method (Savitzky-Golay) was employed. This method provides a closer approximation of local peaks and troughs, and thus reduces the chances that differences between simulated and observed variability are caused by local deviations between the data and the smoothed curves.

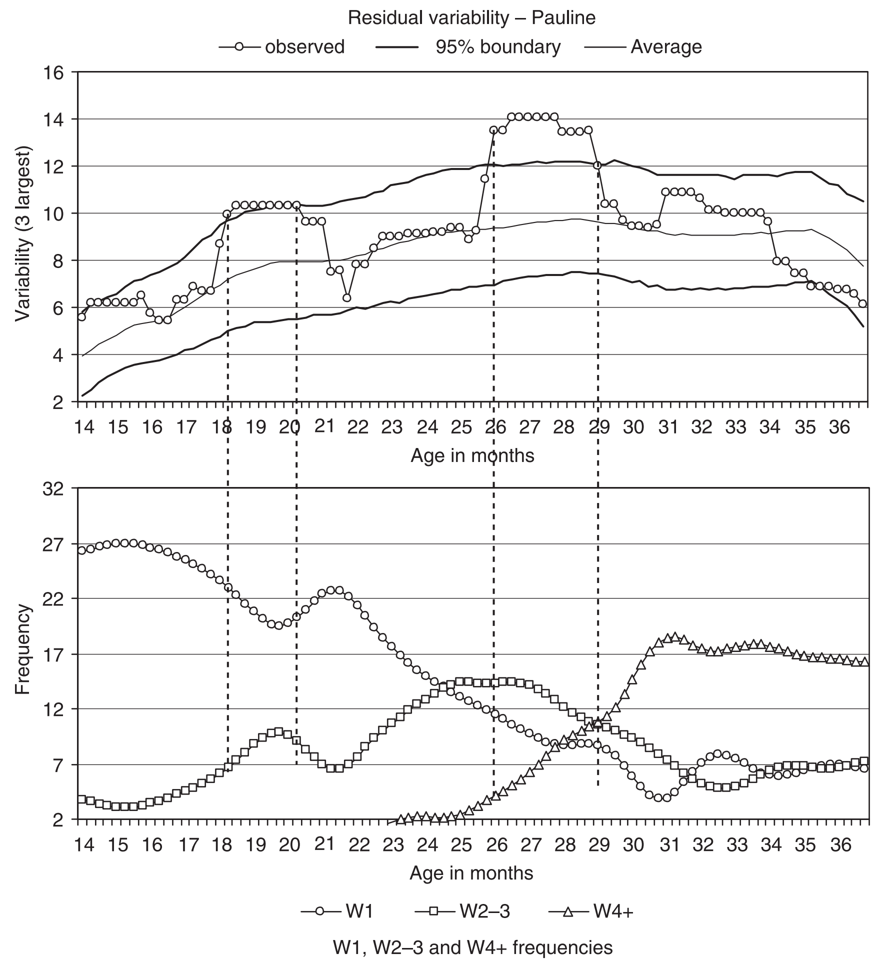


Figure 8 *Residual variability of Pauline's W1, W2–3 and W4+ utterances.*

Figure 1). This is an example of uncertainty about the exact meaning of observed data described by van Dijk and van Geert (2005), who recommend a particular type of sensitivity analysis in such cases. Their method requires that statistical analyses be carried out for 'best case' and 'worst case' scenarios. For instance, since there exists uncertainty as to whether a particular W3 utterance observed at a particular time is actually produced by the combinatorial generator, we must try a scenario in which all W3 utterances are produced by it (our default option, which is likely to produce an over-estimation of the number of utterances produced by the combinatorial generator) and alternative scenarios that are likely to produce under-estimations.

Thus, the statistical analyses of variability were repeated with two alternative groupings, namely (W1, W2, W3+) and (W1–2, W3–4, W5+) (for simplicity, the analyses were confined to residual variability). The first grouping is still consistent with the assumed underlying generators, whereas the second is not (or is to a considerably lesser extent). As predicted, the first, consistent, grouping conserves the temporary dip in W2 utterances and corresponding rise in W1, accompanied by the variability peak. It also conserves the peak in variability in the W3+ utterances, although the peak is somewhat shifted in comparison with the W4+ grouping. With the inconsistent grouping, the temporary regression in the first and second group (W1–2 and W3–4) disappears. In addition, the variability peaks are considerably lower (barely touching or just exceeding the 95% boundary). In summary, the clearest pattern – distinct growth curves and variability peaks – appears with the default grouping (W1, W2–3, W4+) and the consistent alternative (W1, W2, W3+).

A second check concerns the assumption that the late W1 and W2–3 utterances are in fact produced by the syntactic generator and thus must be corrected downward to provide an adequate image of the productivity of the underlying generators. The smoothed curves based

Table 4 *P-values for Benjamin's variability data*

		Longest interval	Second longest interval
p-values within-session variability Benjamin	length	0.034	*n.a.*
	sum of interval	0.038	*n.a.*
	average	0.127	*n.a.*
p-values between-session variability Benjamin	length	0.004	*n.a.*
	sum of interval	0.001	*n.a.*
	average	$p < .001$	*n.a.*
p-values residual variability Benjamin	length	0.001	0.03
	sum of interval	$p < .001$	0.034
	average	0.005	0.029

Note: The *p*-values represent the probabilities that the null hypothesis model produces intervals above the 95% boundary that are as long as or longer than the longest and second longest of such intervals as obtained with the data; since Benjamin's data cover a significantly shorter time span than those of Pauline, *p*-values for combinations and sums of intervals were not calculated; *n.a.* means 'not applicable'.

on this latter assumption were also used in the dynamic growth modeling analysis. To test for variability under this assumption, the observed W1 and W2–3 frequencies must be also be corrected, namely by subtracting (W1 and W2–3) and adding (W4+), the difference from the default smoothed curves with the corrected ones. The corrected data sets produce variability peaks comparable to those found with the uncorrected ones.

In summary, although the relationship between the types of utterances and the presumed underlying generators is probabilistic and also changing across development, the grouping of utterances based on this presumed relationship is sufficiently substantial to warrant the conclusions about two significant peaks in variability.

Statistical test of variability – Benjamin

Within- and between-session variability

Given the shorter and later time span covered by Benjamin's observations, the prediction refers to only one W4+ peak. Figure 9 shows that between-session variability shows a significant peak at the time of maximal growth in W4+ utterances. Within-session variability peaks at the time when W4+ utterance use seems to stabilize, although the use of W4+ utterances is considerably less stable in Benjamin's than in Pauline's case. The temporal order of the within- and between-session variability peaks is similar to that of Pauline.

Table 4 shows that the probability of the null hypothesis model producing an interval of values exceeding the 95% interval is as long as or longer than the observed interval of 0.004 for between-session variability and 0.034 for within-session variability. Recall that statistical significance of the peak in between-session variability around the time of rapid growth in W4+ sentences cannot be ascribed to the rapid growth only, since the null hypothesis model has the same growth pattern as the data. Table 4 shows comparable *p*-values with regard to the total magnitude of the variability peak. If the average magnitude of the peak is considered (row 'average'), within-session variability does not produce significant results ($p = .127$), which is probably due to the fact that the simulated averages can be based on peak intervals of very little length.

Residual variability – Benjamin

The observation that the within- and between-session peaks are further separated than in Pauline's case corresponds with the finding of two adjacent peaks in the current residual variability case (see Figure 10).

Note that the peak corresponding with the between-session variability is considerably greater than that corresponding with the within-session variability. Again, this is not an artifact of the fact that it coincides with the time of rapid growth in W4+ utterances, since the source of variability due to rapid growth is the same in the data and in the null hypothesis model. Since there are two adjacent peaks in the data (corresponding with the within- and between-session cases, respectively), the null hypothesis model was also tested for two peaks. Table 4 (see above) shows that for all criteria, the difference between the observed peaks and the highest peaks based on chance alone is statistically significant (overall, the *p*-values are below the .05 boundary).

Summary of results

For Pauline's as for Benjamin's data, all tests converge on the conclusion that the increase in variability is a real phenomenon and not an accidental artifact of the accelerations and decelerations of the growth of the sentence patterns. Two peaks of variability were found in Pauline's data. One coincided with the local regression of W2–3 sentences. The second peak coincided with the acceleration of W4+ sentences. Benjamin's period of observation is too short to cover both the W2–3 and W4+ transitions, but confirms the peak around the W4+ transition. The difference with Pauline is that Benjamin's within- and between-session variability peaks are further apart and thus form two separate peaks, although, similar to Pauline's data, between-session variability occurs earlier than within-session variability. Finally, the results are relatively insensitive to the way the utterances are grouped or corrected if the grouping or correction is consistent with the model of the underlying hypothetical generators.

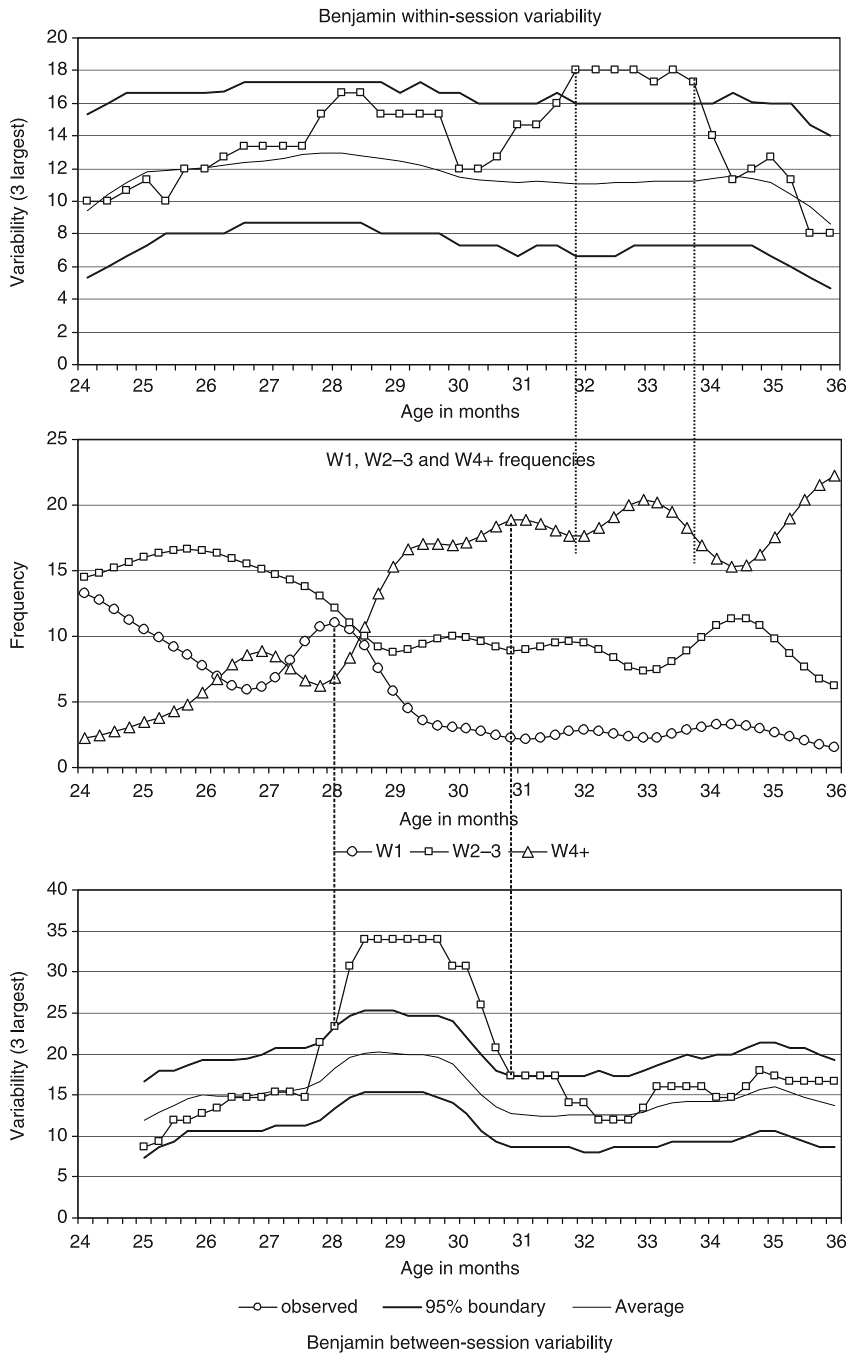


Figure 9 *Total within- and between-session variability of Benjamin's W1, W2–3 and W4+ utterances, compared with smoothed frequency curves.*

Additional support concerning the two critical transition points

To specify which aspects of grammatical development could be responsible for the increase in utterance length and variability in the two transition periods, further analyses were conducted on the linguistic structures of the children's utterances.

With respect to the first transition, we analyzed the W2–3 utterance structures in Pauline's data from 18 to 21 months (Table 5).

Table 5 *Frequencies of the different types of W2–3 utterances in Pauline's corpus, for transition period 18A–21B: for each type, total number (tot nb) and proportion relative to W2–3 utterances (% tot)*

	tot nb	% tot
All W2–3 utterances	**126**	**1.00**
Filler/determiner + noun	**54**	**0.43**
Filler	44	0.35
Determiner	10	0.08
Copula + other word(s)	**17**	**0.13**
Other W2–3 utterances		
Formulaic expression	20	0.16
Interjection + interjection	13	0.10
Interjection + noun	3	0.02
Filler + other	17	0.14
Noun + verb	2	0.02

In order to describe the transition from a linguistic point of view, we looked for utterances contributing to the appearance of the noun clause and to the appearance of the verb clause. Utterances contributing to the appearance of the noun clause corresponded to the 'filler/determiner plus noun' configuration, which dominated the child's production (43% of the W2–3 utterances). A large majority of the 'filler/determiner plus noun' occurrences consisted of the use of a filler with a noun (e.g. */a/ bébé*, '/filler/ baby'), while some infrequent cases consisted of the use of a determiner with a noun (e.g. *le bébé*, 'the baby'). Utterances contributing to the appearance of the verb clause were represented by the 'copula plus other word(s)' configuration, which was relatively frequent (13% of the W2–3 utterances) and generally corresponded to a presentational or attributive construction (e.g. *est pied*, 'is foot'; *est à moi*, 'is to me'). These two configurations, 'filler/

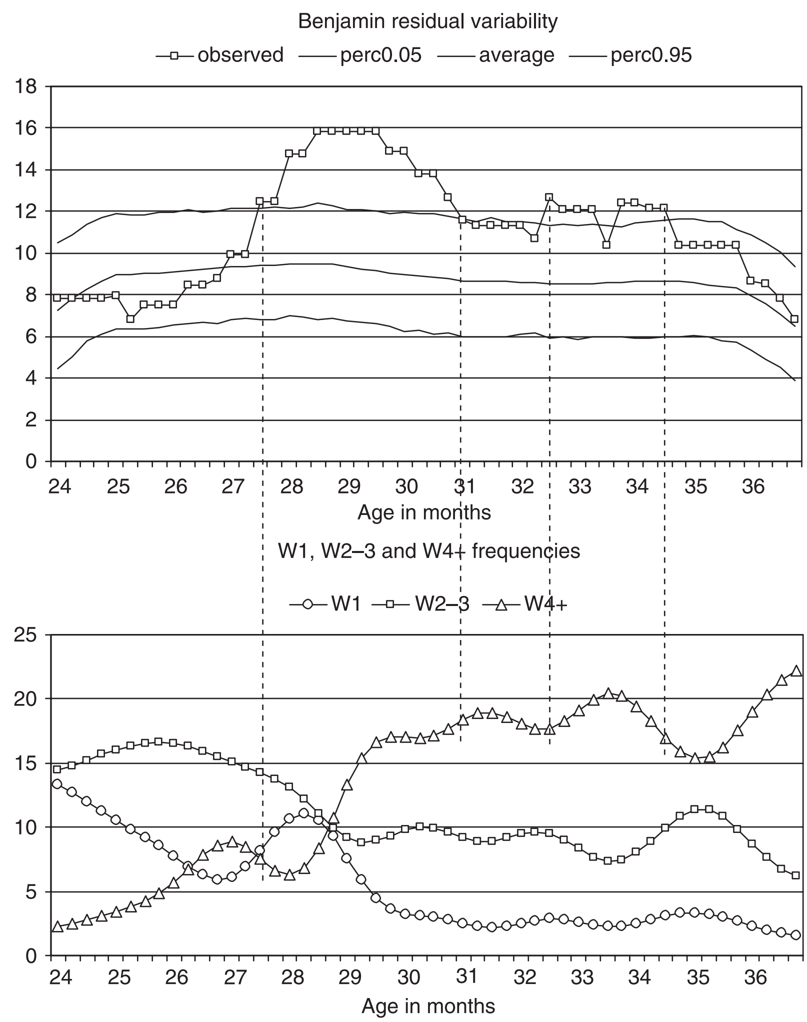


Figure 10 *Residual variability of Benjamin's W1, W2–3 and W4+ utterances.*

Table 6 *Frequencies of the different types of W4+ utterances (simple and complex) in Pauline's corpus, for transition periods 27A–28B and 29A–31B: for each type, total number (tot nb), proportion relative to W4+ utterances (% tot), proportion of correct (% cor) and incorrect (% incor) utterances relative to W4+ utterances*

	Sessions 27A–28B				Sessions 29A–31B			
	tot nb	% tot	% cor	% incor	tot nb	% tot	% cor	% incor
All W4+ utterances	**68**	**1.00**	**0.45**	**0.55**	**189**	**1.00**	**0.90**	**0.10**
Simple utterances	**48**	**0.71**	**0.41**	**0.30**	**136**	**0.72**	**0.67**	**0.05**
Without verb, formulaic	14	0.21	0.21	0.00	22	0.12	0.12	0.00
Identificational	2	0.03	0.00	0.03	21	0.11	0.11	0.00
Attributive	12	0.18	0.15	0.03	23	0.12	0.11	0.01
Imperative, intransitive	7	0.10	0.01	0.09	20	0.11	0.10	0.01
Transitive	13	0.19	0.04	0.15	50	0.26	0.23	0.03
Complex utterances	**20**	**0.29**	**0.04**	**0.25**	**53**	**0.28**	**0.23**	**0.05**
Infinitive clause	12	0.18	0.01	0.17	19	0.10	0.08	0.02
Relative clause	1	0.01	0.01	0.00	5	0.03	0.03	0.00
Other complex utterances	7	0.10	0.02	0.08	29	0.15	0.12	0.03

determiner plus noun' and 'copula plus other word(s)', contrasted with the other kinds of W2–3 utterances found in the child's production (e.g. formulaic expressions, repeated interjections, combinations of interjection plus noun, combinations of filler plus other word). As expected with respect to the first transition period, they are likely to show the emergence of a simple combinatorial stage of language, characterized by proto-grammatical forms of the noun and verb clauses.

With respect to the second transition, we analyzed the W4+ utterance structures in Pauline's and Benjamin's data for the periods of between-session and within-session variability (see Table 6 for Pauline's data).

Following Tomasello (2003), we classified utterances as simple (one main clause) or complex (including one or more subordinated or coordinated clauses). Simple utterances were classified into constructions varying in degrees of syntactic sophistication. The more sophisticated simple constructions are transitive constructions, which require two or more arguments (unlike imperative or intransitive constructions, for instance). In both children, transitive constructions (e.g. *je veux une feuille*, 'I want a sheet'; *on va mettre la couche*, 'we are going to put on the diaper'; *tu me les coupes*, 'you cut them for me') constitute the most frequent type of simple W4+ utterances during the period of between-session as well as within-session variability. Their frequency increased from the first to the second period (in Pauline, 19% and 26% of all W4+ utterances; in Benjamin, 24% and 33%). This analysis suggests that the increasing production of transitive constructions is a major phenomenon contributing to the increase in utterance length and variability during these periods. Another phenomenon likely to contribute to children's utterance lengthening is the appearance of complex sentences, which formed around a quarter of all W4+ utterances during the periods under study. Proportions of complex utterances did not change markedly from the first to the second period (in Pauline, 29% and 28% of all W4+ utterances; in Benjamin, 23% and 21%). They mostly include infinitival complement constructions (e.g. *je veux faire un dessin*, 'I want to make a drawing'), which were the most frequent complex constructions in both periods and in both children. They also include relative clause constructions, adverbial clauses (e.g. *ben, parce que la porte, elle était ouverte*, 'why, because the door, it was open'), juxtaposed or coordinated clause constructions and multi-clause utterances, which were generally produced more frequently in the second period. Finally, a last phenomenon contributing to utterance lengthening during these periods is that both children increasingly produced dislocated constructions, i.e. specific constructions very frequently used in oral French and involving additional pronouns (e.g. *Claire, elle en veut pas*, 'Claire, she does not want it').

Among these three linguistic phenomena contributing to utterance lengthening in children, two at least – the increasing production of transitive constructions and the emergence of complex sentences – clearly reflect a progression in syntactic development and show the appearance of an adult-like syntactic level of language. However, various aspects of the children's productions indicate that we capture here the beginning of the process. In particular, a number of the children's productions were incorrect constructions, such as incomplete transitive constructions in which the subject or object was lacking (*veux lire ça*, 'want to read this' or *je vais pas faire*, 'I am not going to do'), or incomplete infinitive constructions in which the main clause was lacking. However, these incorrect constructions strikingly decreased in frequency in the second period.

General discussion and conclusion

This study addresses the nature and shape of developmental changes in a basic process of early language acquisition, namely children's progression in utterance length, which is related to development in language proficiency and grammar. Based on congruent views in constructivist approaches to language acquisition and in dynamic systems theory, we analyzed the process of increasing utterance length with the purpose of determining possible critical points for the development of grammar.

Evidence for the relationship between changes in utterance length and grammatical development

The central issue addressed by the present study concerns the nature and meaning of changes in utterance length and their relations to grammatical development. Four convergent sets of results shed light on this question and can be summarized as follows. First, the developmental patterns of W1, W2–3 and W4+ utterances were strikingly contrasted. The W1 utterances dominated production until the end of the second year and showed a declining trajectory. The W2–3 utterances presented an inverted U-curve. Finally, the W4+ utterances showed a later and relatively rapid S-shaped increase and began to dominate from the 28th month on in both children. We conceived of the three types of utterances as indicators (under constraints specified, among others, in Figure 1) of temporal attractor states of the developing utterance production system. It is hypothesized that these attractor states show the passage from a dominant holophrastic (non-combinatorial) stage of language to a simple combinatorial stage, and finally to a more sophisticated stage of grammar with syntactic categories and structures resembling those of mature speakers.

Second, a further step in the analysis consisted of the building of a dynamic growth model of the three presumed generators. The simulation results converged on the smoothed data curves and supported the dynamic growth model hypothesis, that simpler forms have a supportive and/or a conditional relationship with the more complex forms, whereas there is a competitive relationship from the more complex to the simpler forms (cf. van Geert, 1991, 1994, 2003).

The third set of results comes from the analysis of variability. We proposed that critical points, which refer to qualitative changes in underlying developmental processes, are marked by anomalies such as temporary regressions and fluctuations, i.e. increasing intra-individual variability. Two points of statistically significant concentrations of augmented variability were shown to occur in the children's data, in the form of separate sets of temporal increases in variability. The first peak was preceded by rapid growth in W2–3 utterances and coincided with the temporary regression of these utterances (remember, however, that Benjamin's data did not cover this age range). The second peak occurred during the period of maximal growth of the W4+ utterances for the two children. In Pauline, the between-session and residual variability peak occurred around 27–28 months, which coincides with the rapid growth phase (this variability peak is significantly higher than would be expected on the basis of the rapid growth alone). The within-session variability peak succeeds the between-session peak and coincides with the phase of relative consolidation of the W4+ utterances, around 29–31 months. A similar timing and order of the variability peaks are found in Benjamin's data.

Finally, qualitative analyses of utterance structures suggest that the first transition period is characterized by the emergence of proto-grammatical forms of noun and verb phrases, and the second by the emergence of syntactic and complex constructions. This view fits in with results from previous studies. Analyses focusing on how Pauline learned to use obligatory pre-nominal determiners and obligatory pre-verbal auxiliaries or modals showed grammatical explosions occurring between 27 and 29 months for the noun grammaticization process, and between 28 and 30 months for the verb grammaticization process (Bassano, 2000; Bassano *et al.*, 2004). Such a coincidence in timings concerning more specific aspects of grammatical development than utterance length justifies the view that the rapid consolidation of W4+ utterances reflects the passage to a more complex syntactic stage.

All these results converged to provide evidence that there exist two critical points in the gradual development of grammar in children, corresponding with hypothesized transitions from the holophrastic to the simple combinatorial stage of language, and from the simple combinatorial to the syntactic stage. The hypotheses fit the constructivist approaches that argue for an early combinatorial stage followed by the development of a more abstract syntax from around 30 months of age.

Critical challenges of the two-transitions model

To begin with, the constructivist view that children's abstract linguistic knowledge develops relatively late is questioned by studies of infant processing of language and language-like stimuli. This issue was discussed extensively in Naigles' (2002) review which shows that even prelinguistic infants find abstract patterns in speech-like stimuli in statistical learning experiments, and, more importantly, that young children are, to some extent, sensitive to certain morphosyntactic English patterns in preferential-looking experiments. Such discrepancies in assessing children's grammatical performances can be related to various phenomena in language development, such as differences between comprehension and production, as well as differences between the processes of acquiring forms without meaning vs. integrating forms with meaning (Naigles, 2002).

We assume that, if sensitivity to the abstract patterns of language indeed occurs precociously, children will have a cumulative experience with such aspects from an early age, a viewpoint which is compatible with that of constructivist approaches. Such accumulation of experiences is likely to be continuous and relatively linear. However, does this continuous, early-emerging experience with abstract, grammatical properties of language conflict with our finding of two major discontinuities? The answer is that in complex dynamic systems such as the developing language, gradual changes in important underlying affecting variables (identifiable as control parameters) are likely to result in non-linear, discontinuous changes in the structural organization of the system (expressed by the system's order parameters). If cumulative experience with abstract linguistic properties is viewed as one of the – potentially many – control parameters that affect the acquisition of language, its postulated continuous change is far from incompatible with non-linear changes and discontinuities in the overall structure of the language production system.

Another objection against our model concerns the relationship between W1, W2–3 and W4+ utterances and the holophrastic, combinatorial and syntactic generators, respectively. To begin with, does this grouping of utterances relate in a meaningful way to the course of grammatical development? If the classification of utterances into these three groups does not sufficiently correspond, in the probabilistic sense, with an underlying division in qualitatively distinct generators, there is also no reason to expect that this particular classification will result in the occurrence of statistically significant phenomena, such as variability peaks and regressions. On the other hand, if there were only one continuously developing generator from the beginning – and this is irrespective of the exact nature of that generator – there would also be no reason to expect any of the variability peaks actually found in our study. We have also tested, and rejected, the possibility that alternative groupings show the predicted peaks. Finally, a preliminary analysis of intra-individual variability in two Dutch children, described in van Dijk and van Geert (2005), showed that there is evidence of two statistically significant peaks at about the same point as with the French-speaking children.

Assuming that the validity of the W1, W2–3 and W4+ grouping can indeed be accepted on the basis of the discontinuities it reveals, we are still faced with the question of whether these discontinuities correspond with the hypothesized holophrastic, combinatorial and syntactic generators. The truth is that we have only indirect justifications for the contention that they indeed do so. Two types of justification have been given in this article. The first relates to the fact that a three-generator model correctly predicts the observed discontinuities, whereas a single-generator hypothesis does not lead to such a prediction. It is of course true that the finding of two discontinuities does not in itself prove anything about the nature of the underlying generators. However, a second type of justification given in the present article, relating to the link between the children's progression in utterance length and their progression in grammatical development, has provided further support for our hypothesis of a succession of holophrastic, combinatorial and syntactic generators.

To the latter we can add a more comprehensive but necessarily rather sketchy justification which starts from the major properties of language. These major properties are, first, that language entails a connection between sound (form) and meaning, second, that it consists of combinations of a finite set of elements (e.g. words into utterances) and, third, that this combination is recursive and governed by a set of rules that are to a considerable extent language specific. If language emerges through self-organization of a great many influences and factors, it should not appear unlikely that these major properties correspond with – transient – states in the self-organizational process. Thus, naturally beginning with sound–meaning couplings (which in our description take the form of basically one-word utterances) the system is apt to rapidly settle into a transient stage of elementary combinatorics (which in our description is approximately taking the form of two- to three-word utterances). This phase is followed by an equally rapid transition to a stage where the possibilities of combination through recursive embedding develop in close association with the child's discovery of the ways in which the ambient language makes this particular combination come true. This latter stage is probably marked by – but is certainly not identical to – the rapid increase of more-word utterances that explicitly express syntactic properties.

This broad interpretation of the potential meaning of the discontinuities that are presented in this article is, by necessity, speculative and preliminary. We can only suggest that further evidence be collected which addresses, among others, languages other than French, and which also entails a more thorough analysis of the qualitative linguistic changes that accompany the hypothesized transitions suggested by the quantitative data.

Acknowledgements

This research was supported by a grant from the '*Cognitique*' program (French Ministry for Research). We would like to thank the children and parents who took part in our study, as well as the collaborators who contributed to the data collection and transcription.

References

Armon-Lotem, S., & Berman, R.A. (2003). The emergence of grammar: early verbs and beyond. *Journal of Child Language*, **30**, 845–877.

Barrett, M. (1982). The holophrastic hypothesis: conceptual and empirical issues. *Cognition*, **11**, 46–76.

Bassano, D. (1996). Functional and formal constraints on the emergence of epistemic modality: a longitudinal study on French. *First Language*, **16**, 77–113.

Bassano, D. (2000). Early development of nouns and verbs in French: exploring the interface between lexicon and grammar. *Journal of Child Language*, **27**, 521–559.

Bassano, D., Maillochon I., & Eme, E. (1998). Developmental changes and variability in early lexicon: a study of French children's naturalistic productions. *Journal of Child Language*, **25**, 493–531.

Bassano, D., Eme, P.E., & Champaud, C. (2005). A naturalistic study of early lexical development: general processes and inter-individual variations in French children. *First Language*, **25** (1), 67–101.

Bassano, D., Laaha, S., Maillochon, I., & Dressler, W.U. (2004). Early acquisition of verb grammar and lexical development: evidence from periphrastic constructions in French and Austrian German. *First Language*, **24** (1), 33–70.

Bates E., & Goodman J.C. (1999). On the emergence of grammar from the lexicon. In B. MacWhinney (Ed.), *The emergence of language* (pp. 29–79). Hillsdale, NJ: Lawrence Erlbaum Associates.

Berman, R.A. (1986). A step by step model of language learning. In I. Levin (Ed.), *Stage and structure: Re-opening the debate* (pp. 191–219). Norwood, NJ: Ablex.

Blake, J., Quartaro, G., & Onorati, S. (1993). Evaluating quantitative measures of grammatical complexity in spontaneous speech samples. *Journal of Child Language*, **20** (1), 139–152.

Bornstein, M.H., Painter, K.M., & Park, J. (2002). Naturalistic language sampling in typically developing children. *Journal of Child Language*, **29** (3), 687–699.

Borsboom, D., Mellenbergh, G.J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, **110** (2), 203–219.

Borsboom, D., Mellenbergh, G.J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, **111** (4), 1061–1071.

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

De Weerth, C., van Geert, P., & Hoitink, H. (1999). Intra-individual variability in infant behavior. *Developmental Psychology*, **35** (4), 1102–1112.

Elman, J.L., Bates, E., Johnson, M.H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Emde R.N., & Harmon, R.J. (1984). *Continuities and discontinuities in development*. New York: Plenum Press.

Fischer, K.W., & Bidell, T.R. (2006). Dynamic development of action, thought, and emotion. In W. Damon & R.M. Lerner (Eds.), *Theoretical models of human development: Handbook of child psychology* (6th edn., Vol. 1, pp. 313–399). New York: Wiley.

Härdle, W. (1991). *Smoothing techniques (with implementation in S)*. New York: Springer Verlag.

Hickey, T. (1991). Mean length of utterance and the acquisition of Irish. *Journal of Child Language*, **18** (3), 553–569.

Hood, G. (2004). Poptools [Computer software]. Pest Animal Control Co-operative Research Center (CSIRO), Canberra, Australia.

Jonhston, J.R. (2001). An alternate MLU calculation: magnitude and variability of effects. *Journal of Speech, Language, and Hearing Research*, **44** (1), 156–164.

Klee, T., & Fitzgerald, M.D. (1985). The relation between grammatical development and mean length of utterance in morphemes. *Journal of Child Language*, **12** (2), 251–269.

Lieven, E.V. (1997). Variation in a cross-linguistic context. In D. Slobin (Ed.), *The cross-linguistic study of language acquisition* (Vol. 5, pp. 199–263). Mahwah, NJ: Erlbaum.

Lieven, E.V., Pine, J.M., & Baldwin, W. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, **24**, 197–219.

Lord, F.M., & Novick, R. (1968). *Statistical theories of mental testscores*. Reading, MA: Addison-Wesley.

MacWhinney, B. (Ed.) (1999). *The emergence of language*. Hillsdale, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk, Vol. 1: Transcription format and programs* (3rd edn.). Mahwah, NJ: Lawrence Erlbaum Associates.

Manly, B.F. (1997). *Randomization, bootstrap and Monte Carlo methods in biology* (2nd edn.). Boca Raton, FL: Chapman and Hall.

Marchman, V., & Bates, E. (1994). Continuity in lexical and morphological development: a test of the critical mass hypothesis. *Journal of Child Language*, **21**, 339–366.

Miller, J.F., & Chapman, R.S. (1981). The relation between age and mean length of utterance in morphemes. *Journal of Speech and Hearing Research*, **24** (2), 154–161.

Naigles, R.L. (2002). Form is easy, meaning is hard: resolving a paradox in early child language. *Cognition*, **86**, 157–199.

Parisse, C., & Le Normand, M.T. (2000). How children build their morphosyntax: the case of French. *Journal of Child Language*, **27** (2), 267–292.

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

Pizutto, E., & Caselli, M.C. (1994). The acquisition of Italian verb morphology in a cross-linguistic perspective. In Y. Levy (Ed.), *Other children, other languages*. Hillsdale, NJ: Erlbaum.

Rollins, P.R., Snow, C.E., & Willett, J.B. (1996). Predictors of MLU: semantic and morphological developments. *First Language*, **16** (47, Pt 2), 243–259.

Rondal, J.A., Ghiotto, M., Bredart, S., & Bachelet, J.F. (1987). Age-relation, reliability and grammatical validity of measures of utterance length. *Journal of Child Language*, **14** (3), 433–446.

Ruhland, R., & van Geert, P. (1998). Jumping into syntax: transitions in the development of closed class words. *British Journal of Developmental Psychology*, **16**, 65–95.

Scarborough, H.S., Rescorla, L., Tager Flusberg, H., & Fowler, A.E. (1991). The relation of utterance length to grammatical complexity in normal and language-disordered groups. *Applied Psycholinguistics*, **12** (1), 23–45.

Siegler, R.S. (1996). *Emerging minds: The process of change in children's thinking* (2nd edn.). New York: Oxford University Press.

Simonoff, J.S. (1996). *Smoothing methods in statistics*. New York: Springer-Verlag.

Thelen, E., & Smith, L.B. (1998). Dynamic systems theories. In W. Damon & R. Lerner (Eds.), *Handbook of child psychology* (pp. 563–634). New York: Wiley.

Thordardottir, E.T., & Ellis Weismer, S. (1998). Mean length of utterance and other language sample measures in early Icelandic. *First Language*, **18** (52, Pt 1), 1–32.

Tomasello, M. (1992). *First verbs: A case study in early grammatical development*. New York: Cambridge University Press.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, **74**, 209–253.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M., & Brooks, P.J. (1999). Early syntactic development: a construction grammar approach. In M. Barrett (Ed.), *The development of language* (pp. 161–190). Hove: Psychology Press.

Traub, R.E. (1994). *Reliability for the social sciences: Theory and applications*. Thousand Oaks, CA: Sage.

Van der Linden, W.J., & Hambleton, R.K. (1997). *Handbook of modern item response theory*. New York: Springer.

Van der Maas, H.L.J., & Molenaar, P.C.M. (1992). Stagewise cognitive development: an application of catastrophe theory. *Psychological Review*, **99**, 395–417.

Van Dijk, M., de Goede, D., Ruhland, R., & van Geert, P. (2003). Child language cuts capers: a study of intra-individual variability in child language. In M. van Dijk (Ed.), *Child language cuts capers* (pp. 11–30). Groningen: Doctoral Dissertation.

Van Dijk, M., & van Geert, P. (2005). Disentangling behavior in early child development: interpretability of early child language and its effect on utterance length measures. *Infant Behavior and Development*, **28** (2), 99–117.

Van Dijk, M., & van Geert, P. (2007). Wobbles, humps and sudden jumps: a case study of continuity, discontinuity and variability in early language development. *Infant and Child Development*, **16** (1), 7–33.

Van Geert, P. (1991). A dynamic systems model of cognitive and language growth. *Psychological Review*, **98**, 3–53.

Van Geert, P. (1994). *Dynamic systems of development: Change between complexity and chaos*. New York and London: Harvester Wheatsheaf.

Van Geert, P. (1998a). A dynamic systems model of basic developmental mechanisms: Piaget, Vygotsky and beyond. *Psychological Review*, **105** (4), 634–677.

Van Geert, P. (1998b). We almost had a great future behind us: the contribution of non-linear dynamics to developmental-science-in-the-making. *Developmental Science*, **1** (1), 143–159.

Van Geert, P. (2001). Fish, foxes, and talking in the classroom: introducing dynamic systems concepts and approaches. In H.A Bosma & E.S. Kunnen (Eds.), *Identity and emotion: Development through self-organization. Studies in emotion and social interaction* (pp. 64–92). New York: Cambridge University Press.

Van Geert, P. (2003). Dynamic systems approaches and modeling of developmental processes. In J. Valsiner & K.J. Conolly (Eds.), *Handbook of developmental psychology* (pp. 640–672). London: Sage.

Van Geert, P., & van Dijk, M. (2002). Focus on variability: new tools to study intra-individual variability in developmental data. *Journal of Infant Behavior and Development*, **24** (6), 1–35.

Van Geert, P., & van Dijk, M. (2003). Ambiguity in child language: the problem of inter-observer reliability in ambiguous observation data. *First Language*, **23** (3), 259–284.

Veneziano, E., & Sinclair, H. (2000). The changing status of 'filler syllables' on the way to grammatical morphemes. *Journal of Child Language*, **27** (3), 461–500.

Veneziano, E., Sinclair, H., & Berthoud, I. (1990). From one word to two words: repetition patterns on the way to structured speech. *Journal of Child Language*, **17**, 633–650.

Weisstein, E. (1999). *CRC concise encyclopedia of mathematics*. Boca Raton, FL: CRC Press.

Received: 30 January 2006
Accepted: 15 July 2006