Chapter 4

Wobbles, humps and sudden jumps.

A case study of continuity, discontinuity and variability in spatial prepositions.

M. van Dijk and P. van Geert
Submitted for publication

Abstract
The aim of this article is to present a new approach to the analysis of longitudinal language data (time series), and apply this approach to a dataset of early child speech - in particular data on the development of prepositions and their use across contexts. Our approach focuses on a quantitative analysis of continuity versus discontinuity and on the form of, and eventual changes in, the variability in the data. We will apply this new approach to our own dataset of early child language (n=4), and present the results with regard to continuity versus discontinuity in the acquisition of prepositions. In our discussion we will address several fundamental issues concerning (dis)continuity and variability.
4.1 Introduction

4.1.1 Continuous versus discontinuous developmental processes

4.1.1.1 Definitions of (dis)continuity
Developmental change in general and language development in particular can literally take many shapes, ranging from the linear increase via the classical S-shape to a trajectory with discrete developmental steps (stages). One important distinction among these different shapes of development concerns the difference between continuity and discontinuity. The meaning of the terms “continuity” and “discontinuity” differs across studies and disciplines. For instance, while Sternberg and Okagaki (1989) use the term to describe lack of smoothness in growth curves and the lack of constancy of the rank orders of individual scores across time, Ford and Lerner (1992) use it to indicate different ways and causes of development (see Ruhland, 1998). Wimmers (1996) stated that "a [non-equilibrium] discontinuous phase transition [in a complex system] involves an abrupt shift from one stable configuration of behavior to another, due to breaking and (re)making the underlying constraints" [pp. 7]. Van Geert, Savelsbergh and van der Maas (1999) state "we can define a discontinuity in the most general sense by identifying it with a stage transition. That is, if a stage is replaced by another stage, in the sense of consecutive sets of different equilibria, a discontinuity has occurred" [pp. XV]. Ruhland gave the following definition: "there is a qualitative, structural change in the development, in such a manner that there is no structural linkage with earlier behavior, and this change is reflected in the lack of smoothness of a growth curve" [pp. 15].

It is important to note that discontinuity is in principle qualitative (structural) by nature. One stable state (equilibrium) is followed by a qualitatively different stable state. For instance, in the field of motor development, a qualitative discontinuity in the development from crawling to walking can be observed. Crawling and walking are qualitatively different, and once children learn to walk, crawling rapidly extinguishes. According to Ruhland’s definition, this qualitative change should also be reflected at the quantitative level. However, we suspect that even if a new developmental aspect emerges abruptly, the quantitative replacement of the old form by the new may occur in a continuous fashion. On the other hand, the finding of a quantitative discontinuity may indicate a qualitative discontinuity (unless there is clear evidence that it is caused by some independent external factor that suddenly emerges).

4.1.1.2 Examples of (dis)continuity
In order to give the reader an impression of continuous versus discontinuous growth curves, we will now discuss some examples. Figure 4.1 shows four typical continuous growth curves. These are linear increase, exponential increase, asymptotic change and logistic growth (see figure 4.1).
Wobbles, humps and sudden jumps.

Figure 4.1. Examples of linear increase (1A), exponential growth (1B), asymptotic increase (1C) and logistic growth (1D).

The simplest form of change is linear (figure 4.1A), which shows a constant additive increase over time. Exponential increase (figure 4.1B) is characterized by a slow start, and an infinitely increasing steepness. Asymptotic growth (figure 4.1C) appears when the onset of change is very rapid, followed by decrease and the reaching of a maximum. Logistic (or S-shaped) change (Figure 4.1D) is a combination of the exponential and asymptotic change. This curve shows a slow start, followed by an increasing slope, which changes into a decreasing slope, and finally the growth levels out when a maximum is reached.

Figure 4.2. Examples of the discrete step (2A) and the cusp (simplified representation) (2B).

Examples of discontinuous curves are the discrete step (figure 4.2A) and the cusp (figure 4.2B). In these examples, the pattern begins as a horizontal line (discrete step) or a slow increase (cusp). The change occurs suddenly. The dotted line represents the transition from one stage to the second. The
difference between the discrete step and the cusp is that the cusp has an overlap region between these stages and the transitions can occur at different points in time at this overlap region. The cusp has several other characteristics, which will be addressed later on. Important is that in a discontinuity the growing variable jumps from one level to the next without intermediary points. It should be noted that the terms "spurt" and "jump" are often used in relation to discontinuities, while in principle spurts and jumps are not typical of discontinuous change. The reason for this is that the sampling is almost by definition discrete and thus easily misses data that lie along the steep increase. Actually, a continuous curve can look like a discontinuous one. This is especially the case for very steep S-curves. Thus, although a developmental curve might show a definite jump or spurt, this does not automatically imply that the data are discontinuous. This point illustrates the problematic methodology of determining these discontinuities in empirical data.

4.1.2 Catastrophe Theory and the testing for discontinuities

4.1.2.1 The cusp catastrophe as a model for developmental discontinuities
Dynamic systems theory not only provides a convincing theoretical framework for development, but also focuses attention on the more irregular aspects of change. Two central concepts in this approach are the concept of self-organization and the concept of an attractor-value. Discontinuities are of special interest because they may indicate increasing self-organization in the system, that is a spontaneous organization from a lower level to a higher level or order (van Geert, Savelsbergh & van der Maas, 1999). According to catastrophe theory, which can be considered a specific branch of dynamic systems theory, self-organizational processes can be classified into a limited number of characteristic patterns of discontinuous change, depending on the number of fundamental variables that determine the change. As such, catastrophe theory offers concrete models and criteria for discontinuities in developmental processes. The theory distinguishes seven elementary catastrophes. All these catastrophes are discontinuities that are caused by changes in control variables. Of these seven catastrophes, the cusp is the simplest: it is controlled by two variables and there are two equilibrium states. Also, the cusp has been successfully applied to several aspects of development (Hosenfeld, van der Maas & van den Boom, 1997; Jansen & van der Maas, 2002; van der Maas & Molenaar, 1992; van der Maas, 1993; Wimmers, 1996) and social behavior (Tesser & Achee, 1994). The cusp looks like a folded table-cloth (see figure 4.3) with one behavioral variable (z) and two control variables (x and y). Thus, the cusp is a three dimensional model, folded at the front and smooth at the back. The cusp can be fitted directly to data consisting of measurements for x, y and z (Cuspfit; Hartelman, 1996, based on the method of Cobb & Zachs, 1985).
4.1.2.2 Qualitative indicators of the cusp catastrophe: catastrophe flags

A second way of testing the cusp catastrophe model, in addition to direct fitting techniques uses the special characteristics of the cusp model. Catastrophe theory provides eight so-called catastrophe flags to test the presence of a cusp model (Gilmore, 1981). These are: (1) sudden jump (which indicates that the change must be sudden), (2) multimodality (there are two or more modes in the data), (3) inaccessibility (there should be no observations in between modes), (4) hysteresis (the level of the jump depends on the direction of the change), (5) divergence (the jump is critically dependent on the initial conditions), (6) divergence from a linear response (there are large oscillations after a perturbation), (7) critically slowing down (there is a delayed recovery after a perturbation), and (8) anomalous variance (there is increased variability in the vicinity of a jump). If all flags are present in a developmental dataset, we can be sure of a discontinuity in the form of a cusp catastrophe.

4.1.2.3 Catastrophe flags in early language development

Catastrophe flag testing is virtually absent in the field of language development. The only exception is the study of Ruhland (1998, also see Ruhland and van Geert, 1998), who investigated transitions in
the use of function words in early language development. This study analyses the time serial language data of six subjects (ages from around 1;6 till around 3;0) from the Groningen Dutch Corpus (Chiledes database). In longitudinal observations of spontaneous speech, frequency counts of articles, modals and pronouns were made. This selection of function words represents an adequate cross-section of all types of function words. These, as Ruhland claims, can be seen as a measure for syntactic and semantic complexity of the child utterances (Ruhland, 1998). Out of the eight flags, Ruhland found evidence for three flags: sudden jump, multimodality and anomalous variance. However, interindividual differences were large in the sense that not all children showed these flags. After fitting linear, logistic and cusp models to the data (using Cuspfit), the cusp model gave the best fit. Since no control parameters are known in the development of function words, Ruhland used “age” as a control parameter. Ruhland expresses his doubt about whether such a fit is sufficient to reject the continuous models, because the cusp usually fits best to data with a high growth rate. Ruhland concludes that in order to reject continuity, too few children show too few flags.

There are several examples of studies that have employed catastrophe flag testing in other fields of infant development. For instance van der Maas, Raijmakers, Hartelman and Molenaar (1999) applied the flags critically slowing down, sudden jump and anomalous variance to reaction times in the development in analogical reasoning. The evidence for the sudden jump turned out to be not very convincing, but statistically significant anomalous variance and critically slowing down was shown. Wimmers (1996) used the catastrophe flags to detect a discontinuous transition from reaching without grasping to reaching with grasping. This transition occurs in most infants between the ages of 16 and 24 weeks. It was the combination of flags that lead Wimmers to conclude that the development of the perceptual-motor organization constitutes a discontinuous transition. Whereas sudden jump, inaccessibility and bimodality were sufficient to reject the only linear continuous model (and not the logistic model), anomalous variance and other properties related to the stability of the system were strong indicators for discontinuous change.

4.1.2.4 The limitations of cusp catastrophe testing

In principle, only the presence of all catastrophe flags indicates a cusp catastrophe. The examples of "catastrophe flag studies" show that this strict criterion is seldom (if ever) met. This leads us to the question of how many flags, and additional indicators (such as the model fitting and other observations), are necessary and sufficient to indicate discontinuity? It is important to note that not all flags can be observed in longitudinal data. For instance, the test for hysteresis is intrinsically impossible, since it would require the reversing of time. Also, for child behavior that is studied with non-invasive methods, we can also not test divergence, divergence from linear response and critical slowing down, since they can only be tested after perturbation. In the case of early language development, it is hard to conceive of an appropriate –let alone ethically acceptable– perturbation method.

The direct fitting of a cusp catastrophe model is confined to data sets for which estimations of the two control variables are available and thus, by implication, is limited to phenomena that critically depend on only two control variables (although this limitation may be weakened by using latent control
variables that are functions of various observed variables; see Guastello, 1988). Also, in many domains of development (such as early language development), no real control parameters are available.

Finally, the cusp catastrophe plane describes the set of equilibrium levels of a behavior. That is, for each couple of (x,y)-values, the corresponding z-value is a point attractor that the behavior should reach or approach after a sufficiently short time (except for those x-y values in the vicinity of the jump, in which case we should find the divergence “flag”). However, it may be questioned whether the notion of a point attractor applies to the developmental phenomena that we wish to study, such as the spontaneous use of prepositions (which relates to the discussion of “true” scores, see further).

4.1.3 Discontinuity and variability

4.1.3.1 Intra-individual variability
A central finding of many time serial datasets is the occurrence of large intra-individual variability. Intra-individual variability can be defined as: differences in the behavior within the same children, at different points in time. In developmental data, variability is expressed as fluctuations between measurement points, which make the data look capricious. As an example, we refer to the developmental trajectories of function words, found by Ruhland (1998). During all of these trajectories, the intra-individual fluctuations are large. This finding may be problematic for the application of the flag inaccessibility. Given the fluctuations, we may find frequencies in the region that the model specifies as “inaccessible”. However, does this mean that all these variable datasets are inherently continuous? We will address this fundamental question in the discussion section.

Intra-individual variability has largely been neglected until the introduction of dynamic systems theory, about a decade ago. Dynamic systems theory takes a radical departure from the traditional approach to variability. This traditional view is voiced by a strong axiom in psychology called “true score theory”, which systematically considers variability to be the result of measurement error8 (Lord & Novick, 1968; Cronbach, 1960; Nunnally, 1970). This true score theory is closely related to signal-transmission theory. In this theory a signal (analogous to the true score) is being transmitted while "noise" is added to the signal. This theory is often (but incorrectly) understood as if it implies that all (short-term) variability boils down to noise, thus error. This "error hypothesis" is based on the assumption that every psychological measurement is subjected to random measurement error, which is expressed in the variability of repeatedly acquired scores. Dynamic systems theory, however, claims that variability is developmentally meaningful and bears important information about the nature of the developmental process. For instance, Thelen and Smith (1994) claim that in self-organization, the system is attracted to one preferred configuration out of many possible states, but behavioral variability is regarded as an essential precursor. They state: “Variability is revealed when systems are in transition, and when they undergo these shifts, the system is free to explore new and more adaptive associations and

---

8 In modern reliability theory, variables are viewed as random variables and the measurement of a variable is defined as the drawing of a single value from the probability distribution. The true score is defined as the mean or expected value of the random variable (Traub, 1994).
configurations” (pp. 145). Thelen and Smith encourage researchers to investigate the variability in their data.

4.1.3.2 Examples of studies on intra-individual variability

Inspired by this new definition, several researchers have taken up on the study of variability and do as Thelen and Smith suggest: they treat their variability as data, and analyze it. In general, various studies emphasize different aspects of variability. Both Bertenthal (1999) and de Weerth and van Geert (1999, 2002) point to variability as a causal factor for development. Bertenthal discusses the function of (qualitative) variability in the development of crawling in infants. He states: "[..] this variability is not merely a correlate of change but instead a contributor to the change itself" (pp. 105). (see also Bertenthal & Clifton, 1998; Newell & Corcos, 1993). He goes on stating that variability offers flexibility, which drives development following Darwinian principles. Principles of variation and selection make that the most successful behaviors will be repeated and stored more often than the less successful ones. De Weerth and van Geert point at a possible adaptive strategy of variability in early infant emotion-related behavior. Intra-individual variability, they claim, ensures the infant of continuing maternal care: "[..] mother and infant try out new ways of communicating with each other, and also change them over time. [..] [They] tune into each other and influence each other with their moods, attitudes and developing skills etc." (pp. 11). Alibali (1999) found that conceptual change (in the contexts of solving mathematical equations) is a cyclical process, which results in periods with more variability alternated with periods with less variability. In this case, variability is a predictor for the type of strategy change the child is going through. The expectation is that children with low initial variability will generate many new strategies, which causes the variability to increase. At the same time, the children with high initial variability are expected to drop strategies, which causes the variability to decline. In the example of Ruhland (1998), the data were extremely variable. Although within-session variability did not correlate systematically with a developmental transition, between-session variability might still show such a pattern after further analysis. Ruhland states that between-session variability can be considered a developmental characteristic and that knowledge about this phenomenon is important for a more thorough insight in developmental processes.

4.1.3.3 Conclusion: relating (dis)continuity and intra-individual variability

In summary, since the introduction of dynamic systems theory, many current variability-centered studies have appeared in the literature (see for example the special issue on variability in Infant Behavior and Development, issue 25(4), 2002). In van Geert and van Dijk (2002), we proposed to view developmental data as score ranges. Instead of depicting a single line of development, we suggested to include the local (moving) minima and maxima in the graph. The width of this range is informative, since it gives us information on the flexibility and sensitivity of the system to its environment. Where the maximal value represents the maximal performance of a child under optimal circumstances, the minimal value indicates the vulnerability of this performance to other (less optimal) circumstances. This flexibility and vulnerability might be different for different subjects or for the same
subject at a later measurement occasion, i.e. for different phases in development. In spite of its 
importance, these phenomena of intra-individual variability is an underexposed aspect in the 
discussion on continuity versus discontinuity, mainly since there are no methods to study this 
phenomenon in an integrated approach. However, we will defend the position that (dis)continuity and 
variability are inherently intertwined. In order to describe continuity versus discontinuity in the 
trajectory of any developmental variable or “grower”, we should therefore include patterns of 
variability. Consequently, we refine our description of discontinuity as a transition from one variability 
pattern to a different variability pattern, in the sense that there is a sudden change in this variability 
pattern.

4.1.4 Research on prepositions

4.1.4.1 (Dis)continuity and the notion of stages in language development

In this article we will apply the continuity versus discontinuity distinction to the field of early language 
development, more specifically to the early acquisition of prepositions. In our methods we will appoint 
a central position to intra-individual variability. The reason for this is that both (dis)continuity and 
variability remain largely untouched in this branch of developmental psychology. However, these 
issues are related to an important theme in the acquisition literature, namely whether language 
development consists of stages. Language development is often characterized in terms of stages or 
phases. For instance, this characterization is used in the study of rhyme (for instance Fikkert, 1998), 
and the acquisition of grammar (see for instance van Kampen and Wijnen, 2000). As an example, van 
der Stelt and Koopmans-van Beinum (2000) discuss six stages of speech production, from continuous 
phonation in a single breath to the production of meaningful words. They state that a new stage begins 
when children acquire a new ability in the sense of a milestone, (see pp. 113). However, they go on 
saying that this does not mean that sounds that are produced in a previous stage have disappeared. 
The question still remains as to whether these stages are merely descriptive characterizations or 
whether they are true discrete stages (a sequence of periods in time with coherent elements that have 
a certain degree of stability).

Studying discontinuity in language development (as opposed to, for instance, intellectual 
development) has several practical advantages. First, spontaneous language can be quantified easily 
(by means of frequency counts) and, second, the data collection procedure is not invasive. This is 
important since in order to study the developmental process adequately, frequently repeated 
measurements are needed, which disqualifies all invasive procedures.

4.1.4.2 Examples of studies on the growth of prepositions: qualitative and quantitative aspects

The present study aims at providing a process account of the quantitative development of spatial 
prepositions. We do this because we are convinced that it contributes to a fundamental insight about 
how this linguistic category is acquired. In the literature, studies on the developmental trajectory of 
prepositions are predominantly discussed from a qualitative angle, while quantitative aspects remain 
derunderrepresented. As an example, several researchers (e.g. Johnston & Slobin, 1979) focussed on a 
typical pattern in the acquisition of prepositions: across languages children learn "in", "on" and "under", 
Wobbles, humps and sudden jumps.
before they learn "between", "back" and "forth". However, they did not elaborate on the quantities in which they occur in child speech. Furthermore, it is often noted that the earliest prepositions are closely related to actions, and therefore resemble verbs to a certain degree (Clark, 1978; Tomassello, 1987).

There are only a handful of studies that explicitly report quantitative data on the developmental trajectory of prepositions. Examples are Brown (1973), and Stenzel (1996), who showed individual developmental curves. Brown (1973) elaborates on Eve's use of prepositional phrases (age 1;6 to 2:0). He shows the percentage of obligatory contexts of "in" and "on" in Eve's first 12 samples. It should be noted that Brown selected only the instances of "in" and "on" because—as he states—"these were only frequent enough to yield fairly continuous data" (pp. 263). In the first six of these samples in this trajectory, Eve omitted these prepositions more than she supplied them. In these early samples, the curve moves up and down, but from sample 7 onwards, the obliged prepositions level up to between 90 and 100 percent. Analysis of the later data of Eve (after age 2) showed that these levels remain above 90 percent. Brown also mentions that the data from Adam and Sarah resemble the picture described in Eve, but the absolute frequencies of each sample remained too low to acquire continuous data. In summary, Brown described the developmental trajectory of the "obliged prepositions" in terms of a quantitative change: before this change prepositions are mostly omitted with highly fluctuating data, and later at least 90 percent of the obliged contexts were supplied.

Stenzel (1996) revealed a similar jump-wise pattern. In his study on prepositional case in a bilingual child, he reports the longitudinal data of subject Pascal, age 2;4-4;7. At the beginning of this period, prepositions were only used very infrequently. The first impression of the data was that they could be separated into two distinct phases. He states: ".. [I]n the first phase, the absolute number of tokens is very low, and we find some strange distributions. In the second phase, the number of utterances containing prepositions is rather high in some recordings (and zero in others); and the pattern observed in the first phase vanishes." [pp. 1036].

The study of Leikin (1998) also reports quantitative data on the acquisition of locative prepositions across ages, but these are hard to interpret. Leikin reports the scores (sum scores of correct responses) of five age groups of subjects in a preposition-naming task. These age groups (from age 3;0 to 7;2) show an increase in their mean scores, that (in our opinion) can be adequately modeled with a linear model. Leikin concludes that there was a steady improvement in performance with age. However, it should be noted that averaging over groups of subjects often leads to a smoothed trajectory. Moreover, the fact that the data are cross-sectional by nature, does not permit conclusions about individual trajectories, simply because the individual curves were not measured.

4.1.4.3 Research question

The empirical research question that we will address in this article concerns the (dis)continuity of the development of spatial prepositions. More specifically, we shall investigate whether there exists a discontinuous transition from one variability pattern in the use of spatial prepositions to a different variability pattern. We will pay special attention to the question of the empirical criteria that can be used to distinguish discontinuous from continuous developmental patterns.
4.2 The case study

4.2.1 Data collection
In the case study, four subjects (Heleen, Jessica, Berend and Lisa) were followed from around age 1;6 to age 2;6. In the beginning of the study, the subjects were predominantly in the one-word stage, while at the end of the observation period their language showed various early characteristics of the differentiation stage (see for characteristics of the Dutch differentiation stage Gillis and Schaelaekens, 2000). For details on the subjects see table 4.1. All subjects came from middle class families, who lived in suburban neighborhoods in average to large cities in the Middle and North of the Netherlands. The children were raised in a monolingual Dutch environment, with no apparent dialect. The participating families were recruited through a newspaper article, an e-mail posting and informal contacts. It should be noted that one of the subjects (Lisa) is the first author’s daughter. The subjects’ general cognitive development was tested with the Bayley Developmental Scales 2/30 (van der Meulen & Smrkovsky, 1983) a few months before their second birthday. The scores were average to above average.

Table 4.1
Subject characteristics

<table>
<thead>
<tr>
<th>name</th>
<th>ages</th>
<th>number of observations</th>
<th>gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heleen</td>
<td>1;6,4 – 2;5,20</td>
<td>55</td>
<td>Female</td>
</tr>
<tr>
<td>Jessica</td>
<td>1;7,12 – 2;6,18</td>
<td>52</td>
<td>Female</td>
</tr>
<tr>
<td>Berend</td>
<td>1;7,14 – 2;7,13</td>
<td>50</td>
<td>Male</td>
</tr>
<tr>
<td>Lisa</td>
<td>1;4,12 - 2;4,12</td>
<td>48</td>
<td>Female</td>
</tr>
</tbody>
</table>

The study is based on videotaped observations of spontaneous speech under naturalistic circumstances (the child’s home). During these observations, children and parents were free to follow their normal daily routine. There were a few practical restrictions given to the parents’ activities (such as not watching the television, and not having extensive phone conversations). Parents were also asked to stay in the living room as much as possible. In addition to the child, one of the parents was always present during the observations. In the cases of Heleen and Jessica, there was also an observer present, who operated the camera. This observer was highly familiar to the child and parents. She participated in normal conversation, with the specification that she did not initiate conversation, but only responded when addressed by the child or parent. In the case of Berend and Lisa, one parent operated the camera and no additional observers were present. Siblings were
excluded as much as possible, for instance by scheduling the observations when an older sibling was at school or a younger sibling asleep. All observations took place in the morning, when most children are highly productive (Wells, 1985). Each observation lasted 60 minutes. The camera was positioned in a corner of the living room, overviewing much of the living room space. A wide-angle microphone was attached to the camera, which facilitated the recording of all language while the child moved freely across the room. The quality of the recordings turned out to be adequate: relatively few utterances were missed because of noise in the environment (roughly estimated between 5 and 10 percent). However, because the recordings were relatively long (each observation captured a minimum of 100 transcribed utterances but the majority of observations contained over 200 utterances) and the interference was non-systematic, we are confident that this occasional noise did not influence the results of this study. There was a warming-up time of 5 minutes. In practice, the child hardly noticed the camera and did not behave differently with or without the camera.

The measurement design was scheduled in such a way that both long- and short-term variability could be optimally captured. The general format of the design is based on the common two-weekly measurement design (e.g. from the Childes-database samples). This measurement frequency is considered adequate to study developmental changes. In order to study short-term ("day-to-day") variability, we alternated the two weekly observations with six periods in which observations were intensified to daily observations or observations made every other day. These so-called "intensive periods" were equally divided over the total observation period of a year.

For further analysis of the speech samples, in two children (Heleen and Jessica) all child language and all child-directed adult language were transcribed (orthographically) according to Childes conventions (MacWhinney, 1991) by an experienced transcriber (the first author). Part of the initial transcriptions were made by trained graduate students (one student per subject). However, all files were checked by the first author and altered when her notation differed from the initial transcripts. From these transcripts, the utterances with a preposition were selected with LEGro (Language-analysis Excel add-in Groningen, van Geert, 2000), an Excel-macro that can select utterances with different kinds of criteria. We selected utterances that contained a spatial preposition.

The data of the remaining two subjects (Berend and Lisa) were processed by employing a less time-consuming, more direct approach: only the child’s utterances that contained a preposition were transcribed (Childes conventions), and not the utterances without a preposition. This was done by trained graduate students (one student per subject) after intensive training by the first author. Inter-observer reliability was calculated as the positive overlap-ratio on the basis of an utterance-by-utterance comparison of a sample of the total dataset. These were 0.84 (Heleen), 0.88 (Jessica), 0.87 (Berend) and 0.87 (Lisa)\textsuperscript{9}, which is adequate. (For a discussion on these measures and a critical discussion of the concept of reliability see van Dijk and van Geert, 2003, chapter 6 of this thesis.)

\textsuperscript{9} Overlap ratios of Lisa and Berend are based on independent scoring of prepositions in a sample of two observations (see for a more detailed description Van Geert & van Dijk, 2003, see chapter 6). Scoring the case of Heleen and Jessica was not independent; ratios are based on a sample of 12 respectively 10 observations.
4.2.2 Prepositions-in-context

4.2.2.1 An overview of prepositions used in this study
All prepositions that belong to the set of spatial prepositions were selected, but only if the context was spatial. The total set of spatial prepositions consisted of "in", "uit", "op", "af", "voor", "achter", "tussen", "over", "bij", "naar", "onder", "boven", "binnen", "buiten", "door" (approximate translations are in, out, on, off, before/in front of, after/behind, between, over, near (to/at, to, under, above, in/inside, out/outside, through). We counted the total frequency of prepositions that were uttered in a particular spatial context. This means that if the context showed that the child referred to an object in a spatial relationship to another object, that preposition was included. We included contexts that referred to spatial actions that had just occurred, or that still had to happen. For instance, if a child said "in chair" in the context of "I want to sit in my chair", the preposition was also included. We counted all distinct spatial contexts. For instance if a child said "in chair" and "doll in bed", these were counted as two different spatial contexts. However, because we were interested in the child’s ability to label spatial situations, repetitions were excluded. For instance, if a child repeatedly said "in chair, in chair", while the parent did not respond, this was counted as only one spatial preposition-in-context.

The definition of what is considered “the acquisition of prepositions” varies across studies, related to the theoretical angle of the study. For instance, Brown (1973) used the notion of "obliged prepositions". In this definition, utterance content is considered as criterion to determine the obligatory character of the preposition, which is the case if the context is spatial and the child utters both object and subject of the preposition. The consequence is that the earliest prepositions are excluded from this frequency count, since these prepositions are one-word utterances. Thus, the object and subject are absent by definition. The result is an underestimation of prepositions, especially in the earliest stages of language development.

In this study, we adopted a definition of prepositions that is consistent with the emphasis that dynamic systems theory places on context-dependency. Therefore, we focus on the development of "prepositions-in-contexts", which we define as "the number of distinct spatial contexts the infant labels with a preposition".

4.2.2.2 Research question and expectations
The central question in this study is "is the developmental range of variability of these prepositions-in-contexts continuous or discontinuous?". On the one hand, it might be the case that the development of prepositions-in-contexts can be described as a continuous trajectory. In this case there might be a gradual differentiation of prepositions-in-contexts: from a restricted use to a more refined and flexible application of prepositions in a variety of distinct contexts. In this case, the child acquires the new prepositions context-by-context or type-by-type. From the conceptual standpoint, continuity can be expected if the order of acquisition follows the underlying conceptual complexity of prepositions (Clark, 1978; Johnston & Slobin, 1979), that is, if the acquisition of the related non-linguistic spatial knowledge is continuous. From the modular point of view, continuity can also be expected, as the
result of the mapping of spatial concepts between the three main cognitive modules, namely perception, action and language (van Geert, 1986).

On the other hand, it might be the case that there is a discontinuity in the development of prepositions. In this case, the child suddenly discovers the way prepositions can be used to label spatial relations between objects. This discovery might function as a threshold, after which prepositions are used in abundance, while before they were only seldom used, and only in verb-like situations. A discontinuous trajectory is expected if, for instance, the required syntactic rules are acquired, and these rules generalize instantly across types and contexts. Here, the discontinuity stems from the sudden acquisition of a new structural category, namely the syntactical use of prepositions. This may lead to an instantaneous application of the preposition category to a wide collection of spatial contexts. At first impression, the curves and descriptions presented by Brown (1973) and Stenzel (1996) may resemble this sudden change to some extent, since the data showed two distinct phases in the acquisition of prepositions. However, the data are difficult to compare since the prepositions were counted using different concepts.

4.2.2.3 Variability and context

Furthermore, it might be argued that the variability we observe is highly dependent on (linguistic and non-linguistic) context. Some contexts might be better suited to evoke spatial prepositions than others. Variability in the data therefore not only reflects the possible instability of the developing syntactical system, but moreover coincidental situational factors. We agree that not all variability is a direct reflection of development, more precisely, of the processes of stabilization and destabilization of the acquired linguistic category. On the other hand, we must consider that there exists a mutual interaction between the developing infant and the spatial context. This conception is completely parallel to a basic assumption of dynamic systems theory. Thelen and Smith (1994) for instance, argue, with regard to the development of leg-movements in the first year, that "without a context, there is no essence" [pp. 17]. They go on stating that cognitive development "is [equally] modular, heterochronic, context dependent and multidimensional" [pp. 17]. As we mentioned before, the behavioral repertoire of children of this age is filled with spatial activities, and the child is an active agent in the selection and constitution of activities and topics. Thus, the variability is the result of this mutual relation between the developing child and the context. The child is not passively influenced by context, but also selects it and contributes to it.

Therefore, we must not conceive of the development of prepositions as the development from one stable state (of no preposition use) to another stable state (of having acquired prepositions and using them at a constant level of production). Instead, we must consider the fact that the end-state of development is not a "stable" category, but a category that is "dynamically stable" in the sense that the produced prepositions still show a considerably variable range dependent both on subject-specific and on situational factors. This range of fully acquired preposition use should, however, be smaller than the range of preposition use in young children who are still acquiring this linguistic category. We have some indication that this is indeed the case: in a case study children show noticeably larger variability in their use of spatial prepositions than adults. A statistical test based on a resampling procedure
showed that, on average, variability of four infants (followed from age 1;6 to 2;6, among which Heleen) were larger than that of two adult samples (for details see van Geert and van Dijk, 2002). This suggests that at least part of the variability we wish to describe in this article is developmentally determined.

4.2.3 Statistical analysis and results

The first step consisted of an analysis of the individual trajectories, based on bootstrap and resampling procedures (for general discussions of the bootstrap and resampling method, see for instance Good, 1999, Manly, 1997; the procedures were carried out in Microsoft Excel, by means of a statistical add-in, Poptools, Hood, 2001). The question we addressed is: do the individual trajectories show discontinuities in the developing range? Second, we performed a meta-analysis on the four individual trajectories in order to answer the question whether our conclusions with regard to discontinuities in the individual trajectories also applied to the four subjects as a group.

4.2.3.1 The developmental curves of prepositions

In order to get a general impression of the development of prepositions we plotted the trajectories of the four subjects in figure 4.4. On the basis of visual inspection, we aligned subjects to their first major peak in their trajectories. Using this representation, all subjects show their first major “jump” at point zero on the x-axis. The other numbers on the axis represent the number of days between the other observations and this point zero. Visual inspection of the individual trajectories shows that most subjects can be described in terms of two distinct phases. First, they show a period in which prepositions-in-context only occur occasionally, usually the phase before point zero. In this phase, prepositions fluctuate weakly between values lower than 10. As an exception, Lisa’s initial values reach to around 15. Then, further on, all trajectories are characterized by strong fluctuations, which occur after the first major outbursts of prepositions. Inter-individual differences are clearly visible. For instance, the timing of the transition between this first and second phase differs between the four subjects. While Heleen shows a relatively long initial phase, Berend only shows seven values that might be characterized as such. It can be questioned whether Berend actually displays the initial phase we described above. However, since we have no data before measurement point -38, we can only speculate on the existence of a first phase as found in the other subjects.

Thus, solely on the basis of visual inspection, trajectories of these four subjects may be described in terms of two distinct phases. First, there is an initial phase in which there are only a restricted number of prepositions-in-contexts. Secondly, this phase is followed by a more advanced phase in which prepositions are used in many different contexts, but whose usage is highly variable, dependent on context.
Chapter 4

4.2.3.2 Criteria of discontinuity

In order to test whether this "two-distinct-phases" observation is a reflection of an underlying discontinuity, we need a statistical testing model. The concept of discontinuity we adopted in this study is "a transition from one variability pattern, to a different variability pattern, in the sense that there is a sudden change in this variability pattern", a definition that assigns a central position to variability. Therefore, our testing criterion is designed to be sensitive to changes in these variability patterns. The criteria we constructed for these analyses are based on resampling procedures. The question we intend to answer by means of these procedures is "to what extent is a continuous model capable of producing the results of our four subjects?". This continuous model will be estimated on the basis of the observed data. The continuous model will then be used to simulate data sets, each of which will therefore, by definition, be produced by the continuous model. If these simulated models are capable of producing the statistical indicators of discontinuity that we observed in our subjects, the null-hypothesis of underlying continuous development cannot be rejected. In short, we define a discontinuous model as a model that has characteristic properties that a continuous model does not have. If we had evidence that the process is determined by two control variables, we could postulate that the discontinuity must be consistent with the criteria of the cusp catastrophe. However, there is no evidence that the acquisition of prepositions is governed by two known control variables (or any specific number of control variables, for that matter). Also the application of "age" as a control parameter, as used by Ruhland, is questionable. Finally, we have no reasons to believe that the level of preposition use can be described by a point attractor (the equilibrium level specified by the behavioral plane of the catastrophe, see figure 4.3).
Instead, we looked for two indicators that we think are likely to occur in any kind of discontinuity. The first is the existence of two clearly distinct sub-phases or sub-patterns in a trajectory that shows a discontinuity. If a discontinuity occurs, it is likely to show in the form of two distinct patterns, e.g. distinct patterns of data in terms of distinct means, variability or trends. Of course any pattern or trajectory can be divided in two or more sub-patterns. The question is whether or not it is statistically unlikely that the division based on the observed data can also be made on the basis of the simulated data, i.e. the continuous model. The second indicator is the existence of an anomaly in the data at the moment of the discontinuous shift. We expect that such an anomaly can be observed in the form of an unexpectedly large, local peak or “spike” in the production of prepositions. The peak or spike is “unexpectedly large” if the likelihood that it can be explained on the basis of a continuous model is very small.

Since we characterized discontinuity by properties that a null-hypothesis model of continuity cannot produce, it is important to explain how such a null-hypothesis model was estimated for our data. We defined the null-hypothesis model as the simplest possible and most general model of continuous change in both the trend and the variability of the data. The model that best matches these criteria for the four datasets is a simple linear model with a slope and standard deviation of the residuals that are both a linear function of time. The null hypothesis model for each of the four datasets was calculated by first estimating a linear model for the data. We then calculated the distances between the data and the linear model (the distance is the absolute difference between a data point and the value of the model at that point). We then fitted a linear model to the distances, again based on time as the independent variable. For each moment in time, \( t_i \), the null hypothesis model is a normal probability distribution with a mean equal to \( M_{ti} = \alpha + \beta t_i \) and a standard deviation equal to \( D_{ti} = C .(\alpha D + \beta D t_i) \), for \( C \) a fitting constant\(^{10}\). Finally, note that a null-hypothesis model in the form of either a quadratic equation or a transition equation led to only small improvements in the \( R^2 \)-value in two cases with the quadratic equation and to a worsening in three cases with the transition equation. In summary, we opted for the linear model as the simplest and best fitting null-hypothesis model of discontinuous change in both the trend and the variability.

4.2.3.3 The distance criterion: looking for statistically significant sub-patterns

4.2.3.3.1 Method

The first and strictest definition of discontinuity assumes a model of two distinct phases that can be characterized as two distinct stationary series. That is, each time series (i.e. each sub-phase) has its own constant average and its own constant variability (standard deviation). However, since we are dealing with developmental data, we must relax our assumption of stationarity and accept that each

\(^{10}\) If we fit a model to the local variances, i.e. the squared differences or local variances, the resulting model is too much influenced by the heteroscedasticity of the residuals. Fitting a model to the absolute differences, i.e. the local standard deviations, gives better results, but leads to an underestimation of the variability of the resulting null-hypothesis model. In order to obtain the same level of variability in the data as in the null-hypothesis model, the estimated standard deviation must be multiplied by a fitting constant, which is approximately 1.26.
sub-phase will have its own linear increase, both in terms of average and standard deviation. Thus, we will have a discontinuity if each sub-phase is characterized by its own linear regression model and if the differences between those linear regression models is bigger than can be expected if the data were based on a continuous model. If the developmental trajectory of prepositions is discontinuous, we expect that these linear regression lines show a considerable distance around the moment of the transition (see figure 4.5). This distance (both in absolute terms and in terms of variability) is used as the testing criterion in a resampling procedure. In the case of a discontinuous transition between two score-ranges we are able to formulate the following expectations for each subject:

- The absolute distance between the two linear models is larger in the empirical set than in the simulated set based on the continuous model.
- The difference in standard deviations between the two linear models is larger in the empirical set than in the simulated set based on the continuous model.

In addition, we checked whether the time at which a maximal distance is found between the two regression models describing our observed trajectories, corresponds with the time of the discontinuity estimated on the basis of visual inspection.

Figure 4.5. The distance model in which each sub-phase is characterized by its own linear regression model (with its own average and standard deviation).

On the other hand, if the null-hypothesis of continuous development is true, the trajectory of prepositions basically consists of a single linearly increasing trajectory. If this is the case, the observed frequencies with which prepositions occur will randomly fluctuate around this linear path. With such random fluctuation, however, it is not unlikely that some arbitrary divisions of the entire data set in two subsets will result in linear models for those subsets that also show some discontinuity, i.e. an observable difference between the endpoint of the first subset and the beginning of the second. It is possible, therefore, that the distance between the first and second subset in the data is not an indicator of a two-stage process, but merely the result of random fluctuations over an undivided linear increase. In order to find out how likely the latter explanation is, we have to determine the distances that such a random linear model can produce. The critical issue is: what is the probability that the continuous (undivided linear) model with random fluctuations produces a two-stage distance that is of the same magnitude as the two-stage distance found in our data? The smaller the probability, the less
likely it is that the observed distances in our four subjects are an accidental outcome of an undivided developmental trajectory. This probability is determined by means of a resampling procedure, to be more specific a Monte Carlo analysis (Manly, 1997; Poptools, Hood, 2001).

In order to compute the maximal distance in the data, we divided the individual data in two parts, taking all possible divisions (with a minimal length of 4 observations). For instance, if the series consists of 50 simulated preposition values, we begin with a subset consisting of the first four values and a second subset consisting of the remaining 46. We then calculate a linear model for the first subset and a linear model for the second subset. We determine the distance between the two (the difference between the last value of the first linear model and the first value of the second linear model). We store the value of this difference and make a second division (which would consist of the first five observations and the remaining 45). We repeat the procedure until all possible divisions have been made (which in this example, amounts to 43 different divisions). We select the maximum of all the values obtained: this is the maximal distance value for this particular data series. We proceed by generating a new simulated data series, based on the mean and standard deviation provided by our models. We repeat the maximal distance procedure and keep the maximal distance for this set. This whole procedure of generating a simulated series and calculating its maximal distance is repeated many times. Thus, after repeating this procedure 5000 times, we have 5000 maximal distance values, each of which is based on the null hypothesis model (an undivided linear model with random fluctuation). In addition to taking the values of the absolute distances between the two sub-stages, we will also calculate the relative distance, which is the distance between the two substages, divided by the value of the single-stage linear model at the time the discontinuity between the sub-stages occurs. The reason to do this is that a “jump” of, for instance, size 5 for an expected mean value of 10 is, relatively speaking, more important than if that same jump occurred at a time when the expected mean value is 30 (note that we do not need to calculate the relative distances if the absolute distances are already statistically significant). Finally, we count the number of times the simulated continuous model has produced a maximal distance that is as big as or bigger than the maximal distance we have obtained from our observed data series. This number, divided by 5000 (the number of simulations), gives us the p-value of the observed distance under the null hypothesis as specified.

4.2.3.3.2 Results obtained on the basis of the distance criterion applied to all values

Since the analyses were performed on the individual data, we will report the results for each subject separately (see Table 4.2). We begin with the results based on all the values, both observed and simulated (see further for the second option). The column “estimated position” shows that the distance criterion confirms the discontinuity point based on visual inspection in two cases (0 for Heleen and Lisa), but provides very different estimations of the point of maximal discontinuity in two other cases (259 for Berend and 187 for Jessica) respectively. Thus, the fit between the breaking points based on this procedure on the one hand and visual inspection on the other hand is not very good and at least highly ambiguous.
Second, the average position in the time series at which the maximal distance is found in the simulated series comes considerably later than in the observed data series, where the point was fixed at value 0.

Table 4.2

| Estimated position of observed data series, average positions obtained from simulated data series, p-values of simulated maximal distance, simulated maximal relative distance and simulated difference between standard deviations. Estimations are based on all values |
|---------------------------------|-----------------|--------------|--------------|--------------|--------------|
| Estimated position (observed)   | Average position (simulated) | p-value distance | p-value relative distance | p-value standard deviation |
| Berend | 259 | 179 | 0.40 | 0.68 | 0.69 |
| Heleen | 0 | 33 | 0.08 | 0.03 | 0.16 |
| Jessica | 187 | 143 | 0.25 | 0.43 | 0.56 |
| Lisa | 0 | 89 | 0.54 | 0.30 | 0.36 |
| Meta analysis | | | 0.19 | | 0.51 |

Of more interest are the p-values of the maximal distance between the substages obtained with the simulated, continuous series. The p-values are small only for one child, Heleen (0.08). If we take the relative distance as a criterion, only one child (Heleen) has a p-value smaller than 0.05. In none of the observed cases, the difference between the standard deviations of the observed sub-phases is statistically significant.

In order to summarize the results of the individual subjects, we also performed a meta-analysis in order to transform the individual p-values into a combined p-value for all four subjects together. We did this using a (meta)resampling procedure on the resampling results of the distance criterion. The procedure takes the results of the individual resampling procedures and randomly selects one distance per individual out of the 5000 simulated maximal distance values for each individual subject. Then, the average of these 4 values (one per individual) is taken and compared with the average of the observed maximal distances. The results are also represented in Table 4.2 (bottom). Neither the distance nor the standard deviation criterion produce p-values that come even near statistical significance. In summary, the distance method does not allow us to distinguish the distance between the visually estimated sub-models of the observed preposition levels from those based on a continuous model.
4.2.3.3. Results obtained by applying the distance criterion to the maximal performance level

In various publications, Fischer has claimed that it is not possible to observe discontinuities in the developmental data if the data are based on what he calls the “functional level”, which means the level of unsupported performance (see Fischer and Yan, 2002; Fischer and Rose, 1994). If we monitor the subject’s optimal level of performance – which is achieved by giving the subject support – we will discover discontinuities in that optimal level, which are indicators of stage-wise changes in the subject’s performance. Although we cannot provide our subjects with support, we can nevertheless obtain an estimation of their optimal preposition production by simply taking the local maximum of the preposition frequencies. For instance, we take the maximum level for time windows covering about 10 observations (for a description of this technique, see van Geert and van Dijk, 2002, chapter 3 of this thesis). On the other hand, if we draw a linear model through the data, we obtain an estimation of the average preposition production, which is akin to what we might see as the child’s “functional” level. In practice, we estimated the optimal preposition production by taking the so-called Progressive Maximum (see van Geert & van Dijk, 2002), which is the maximum of a time window that expands from the first until the last observation. See figure 4.6 for the resulting Progmax-regmin graph of the four subjects.

We repeated the distance method by estimating sub-phases in the optimal preposition production and compared this with the estimated sub-phases of the optimal preposition levels of the simulated data sets (i.e. the sets based on the continuous model). As can be seen from Table 4.3, the results are very different and are in line with Fischer’s assumption about discontinuities in optimal levels.

To begin with, the estimated breaking points for the sub-phases in the observed data are very close to those obtained on the basis of visual inspection (0, 5, 0 and 0 respectively). The deviation in the case of Heleen’s data is only 5 days. Furthermore, all the p-values for the absolute distances between the sub-phases, are now statistically significant (p≤< 0.05). Note that the p-values of the standard deviations have been omitted, since they make no sense if we take the values of the optimal performance as the basis for calculation. Since all individual p-values are statistically significant, we do not need to carry out a meta-analysis over the four p-values.
Figure 4.6. Moving Progmax-Regmin graph of prepositions-in-context, Heleen, Jessica, Lisa and Berend.
Table 4.3

Estimated position of observed data series, average positions obtained from simulated data series, p-values of simulated maximal distance. Estimations based on local maximum.

<table>
<thead>
<tr>
<th></th>
<th>Estimated position</th>
<th>Average position</th>
<th>p-value distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berend</td>
<td>0</td>
<td>112</td>
<td>0.03</td>
</tr>
<tr>
<td>Heleen</td>
<td>5</td>
<td>-36</td>
<td>0.01</td>
</tr>
<tr>
<td>Jessica</td>
<td>0</td>
<td>163</td>
<td>0.05</td>
</tr>
<tr>
<td>Lisa</td>
<td>0</td>
<td>53</td>
<td>0.02</td>
</tr>
</tbody>
</table>

4.2.3.4.4 Results obtained by applying the distance criterion to the difference between minimal and maximal performance level

By applying the notion of progressive maximum level we were able to demonstrate a discontinuity in the maximal values, i.e. the positive extremes, in the use of spatial prepositions. However in our Research Question section, we announced to “…investigate whether there exists a discontinuous transition from one variability pattern in the use of spatial prepositions to a different variability pattern.” If, by variability pattern we mean the width of the range within which the use of spatial prepositions can vary from observation to observation, the focus on maximal values cannot, by definition, be used to show that such a discontinuity exists, simply because the maximal level is not an index of variability. In our first analysis, we applied the distance criterion to the standard deviations of hypothetical subsections of the data, but did not find a distance between standard deviations that could not be explained by a continuous linear model. However, as an alternative model of variability of preposition use, we may think of the distance between the extremes, i.e. the distance between the maximal and minimal values. Note that those maximal and minimal values are just as reliable as the more central values: we counted only the productive use of such prepositions and left mechanical repetitions, for instance in the form of songs or word games, out of our counts.

As a measure for the maximum performance level, we took the progressive maximum that we described in the preceding section. As a measure for the local minimum, we took the Regressive Minimum (see van Geert and van Dijk, 2002, for a description). The regressive minimum starts with the value of the last observation and then takes the minimum of a progressively widening window that expands from the last up to the first observation.

We determined the progressive maximum and regressive minimum lines of our four time samples and calculated the distance between the minimum and maximum value for each point in time. We then divided the time range of max-min distance in all possible pairs of subsets (see 4.2.3.3.1 for a description of this method) and calculated the difference between the average of the two subsets. We took the maximal difference as the discontinuity criterion and determined at which point in time this
discontinuity occurred (see Table 4.4). It turns out that this discontinuity lies exactly at the estimated point 0 in our four samples. Put differently, this method replicates our eyeball statistics method.

In order to test whether the maximal distance found in our observed series can be explained by a simple linear model, both of the mean value and the standard deviation, we calculated the maximal distance between the min-max lines of 5000 simulated times series based on the linear models calculated for the observed samples (see Method description). The p-values are 0.04, 0.16, 0.07 and 0.03 respectively (see Table 4.4). A meta-analysis over 5000 combined simulated samples yielded an overall p-value of 0.003. These p-values demonstrate that it is unlikely that a continuous linear model can explain the discontinuity in the variability, defined by the distance between maximal and minimal values.

Table 4.4
P-values of the distance method applied to the distance between maximal and minimal values, based on 5000 simulations

<table>
<thead>
<tr>
<th></th>
<th>Estimated position</th>
<th>Average position</th>
<th>p-value distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berend</td>
<td>0</td>
<td>31</td>
<td>0.04</td>
</tr>
<tr>
<td>Heleen</td>
<td>0</td>
<td>-137</td>
<td>0.16</td>
</tr>
<tr>
<td>Jessica</td>
<td>0</td>
<td>8</td>
<td>0.07</td>
</tr>
<tr>
<td>Lisa</td>
<td>0</td>
<td>-37</td>
<td>0.03</td>
</tr>
<tr>
<td>Meta-analysis</td>
<td></td>
<td></td>
<td>0.003</td>
</tr>
</tbody>
</table>

4.2.3.4 The peak model: looking for a statistically significant spike

4.2.3.4.1 Method

We also employed a second definition of discontinuity to test the presence of discontinuities in our data. While the previous criterion conceptualizes discontinuity in terms of two distinct score ranges, this second criterion focuses on the moment of the transition itself. With this criterion, we identify a discontinuous transition if the trajectory shows an *unexpectedly* large peak or spike, at some point in time. This peak might reveal the moment at which the system losess its stability and shifts into a different variability pattern.

We conceptualized a peak as the maximal "relative difference" between an expected and an observed value in the trajectory. First, variability at a specific point in time was defined as the absolute residual from a (single) linear model. Second, these residuals (for all points in time) were divided by the value of the linear model at the corresponding point in time, which yielded the "relative difference" at every point in time. This concept of "relative difference" is important since the degree of variability is strongly related to the central tendency in a distribution. As an illustration, assume a trajectory that shows a simple increase from value 1 to value 50. In this example, a variance of 4 is considered to be larger
When it occurs around an observation with value 2 than when it occurs around value 49, parallel to what was mentioned previously. Therefore, we need to consider the relative difference, in order to be able to compare fluctuations along the trajectory.

As a third step, we took the maximal "relative difference" in the data as our testing criterion and tested it against simulated continuous models with the same distribution characteristics as the data. However, in order to compare the maximal relative variability between data and simulated models reliably, we had to perform a heteroscedasticity-correction. Because of the statistical phenomenon of heteroscedasticity, the early part of the trajectory might show exceptionally large relative peaks. However, these early peaks are actually an artifact of the low model values that are used to compute the relative variance. For instance, in a dataset that consists of only the values of 1 and 0, the occurrence of the value 2 appears as a large peak, since the increase is 100%. This may be problematic in a dataset that can only consist of whole numbers, since each change must be at least one point. Thus, while the absolute variability is in fact very low, the relative variability is overestimated by these low model values (for a further discussion of heteroscedasticity with regard to the analysis of patterns of variability, we refer to van Geert and van Dijk, 2002, chapter 3 of this thesis). As a solution to this problem, we excluded all measurements before the point in time that the model reaches the absolute value of 4. We did this both with the empirical set and the simulated set. We expect that in case of a discontinuity the maximal relative difference is larger in the empirical sets of the four subjects than in the simulated continuous models.

### 4.2.3.4.2 Results obtained on the basis of the peak criterion

Table 4.5

*P*-values of the peak method based on 5000 simulations

<table>
<thead>
<tr>
<th></th>
<th>Estimated position</th>
<th>p-value peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berend</td>
<td>0</td>
<td>0.42</td>
</tr>
<tr>
<td>Heleen</td>
<td>5</td>
<td>0.38</td>
</tr>
<tr>
<td>Jessica</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Lisa</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Meta-analysis</td>
<td></td>
<td>0.01</td>
</tr>
</tbody>
</table>

Overall, the position of the discontinuity estimated with the peak method corresponds with that based on visual inspection (there is a small difference in the case of Heleen). More importantly, the analysis reveals a striking difference between two children with low p-values (Jessica and Lisa) and two with p-values that are indistinguishable from the null-hypothesis model (the linear, continuous model). However, the p-values obtained with the distance method based on all observed values, showed that
Heleen’s data contain evidence for a discontinuity, whereas the others only showed a discontinuity if the calculations were made on the basis of the local maximum. Visual inspection of the data shows a striking difference that is consistent with the current findings. Heleen’s dataset looks like a pattern of step-wise increase, whereas Lisa and Jessica show a linear pattern punctuated by a sudden burst of productivity (note that all automatic repetitions, such as songs or word games, have been removed from the data, which implies that the peaks cannot be explained on the basis of an accidental and relatively meaningless play with prepositions).

The meta-analysis showed that combining the results of the four subjects leads to a p-value of 0.01. However, it should be noted that this value is probably mainly caused by two highly significant subjects (Jessica and Lisa). The other two subjects (Heleen and Berend) also show a large peak in their trajectory, but it is not statistically significantly different from peaks that can occur in continuous linear models with a similar amount of variability as the observed data. In summary, it can be concluded that according to the “peak definition” of discontinuity we have sufficient indications to reject the null-hypothesis of continuous development of prepositions-in-contexts.

4.3 Conclusion and discussion

4.3.1 General summary and conclusion
On the basis of visual inspection we assumed that the development of prepositions-in-contexts of these four subjects shows a characteristic pattern of two phases. First, there is an initial phase in which prepositions are only used in a restricted number of contexts, followed by a second phase in which prepositions are used in many different contexts, the usage of which is however highly variable.

In order to test whether this observation testifies of a discontinuity, we tested the data using two criteria. The first testing criterion, called the “distance criterion”, is based on the distance (both absolute and in standard deviation) between two linear regression lines on the maximal division of two distinct phases in the trajectory. Using this criterion, the null-hypothesis that the empirically found values are produced by an underlying continuous distribution could not be rejected. However, if we repeated the procedure for the optimal levels of preposition production and for the distance between the minimal and maximal levels, we found strong statistical support for the discontinuity hypothesis. The second criterion, based on the presence of an unexpectedly large peak at the moment of transition, showed that the empirical sets show significantly larger peaks than the simulated continuous data sets in only two cases (Lisa and Jessica). In summary, strong support for a discontinuity is found if we take the shifts either in the maximal level of production or in the difference between the maximal and minimal levels of production.

4.3.2 Two central assumptions
It is worthwhile to discuss two central issues that strongly influence the discontinuity discussion. The first issue is that continuity and discontinuity are usually treated as categorically distinct and that a
developmental process is considered to be either the one or the other. The second issue is that when testing for discontinuities, the null hypothesis is assumed to be continuity. These two issues are interrelated and can be held responsible for the fact that many studies on discontinuities in development present inconclusive results instead of uncovering the more subtle characteristics of continuity and discontinuity.

With regard to the first assumption, we will argue that the distinction between continuity and discontinuity amounts to a matter of degree rather than to a categorical distinction. With regard to the second assumption, the null-hypothesis in a discontinuity study is usually formulated as a continuous process of development. Put differently, continuity is the default option. Only if continuity can be rejected, the developmental process can be regarded as discontinuous. The catastrophe flag studies we discussed (Ruhland, 1998; Wimmers, 1996; van der Maas, Raijmakers & Molenaar, 1999) take this position. However, it should be noted that the null-hypothesis might also be reversed and reformulated as "a process is discontinuous unless proven otherwise". This position is at least equally defendable since a discontinuous process can also explain continuous results at the individual level, while a continuous process cannot explain that some children show discontinuities while others do not. Thus, discontinuity is a more likely candidate for the default option than continuity.

4.3.3 The alternative

3.3.1. Discontinuity and continuity as two ends of a continuum

The alternative is to conceive of discontinuity as a collection of characteristics that each describe different aspects of the discontinuity. In this case "pure" continuity would be described as a complete absence of these characteristics, while "pure" discontinuity indicates the presence of all characteristics. These two situations can be described as the two extreme positions on a continuum, between which many intermediate positions are possible. Characterizing a developmental process by describing which aspects of it are continuous and what aspects are discontinuous may provide more information than merely describing this same process in terms of the traditional either/or-definition of discontinuity. This is especially the case regarding the fact that the finding of one of these extremes (namely the discontinuity-extreme) is quite unlikely. It is important to note that this conception of a continuum between continuity and discontinuity can be combined with the original description of the cusp. As can be seen in figure 4.3, the fold that characterizes the cusp "unfolds" itself with the y-dimension. While at the surface of this y-dimension the fold is at its most extreme position, it is absent at the far end of this dimension. This characterizes the bifurcation that is inherent in the cusp catastrophe: while the far end describes only one mode of behavior, there are two modes in the surface plane between which the system might switch. At what position a system functions depends on the values of the control parameters of the system. Thus, dependent on the interaction between a developing individual and its environment, the resulting trajectory can be described in terms of its position on the y-dimension, which can be considered as the proposed (dis)continuity continuum. Although we doubt whether a discontinuity necessarily presents itself as a cusp, we stress that this idea of a continuum between pure continuity and discontinuity is not contradictory with catastrophe
theory. However, our new definition provides room for additional criteria to define discontinuity, since in most cases the precise number of control parameters is undefined, whereas in case of the cusp it is confined to two. It should also be noted that, in terms of observable variables, a continuous model can be a special case of a discontinuous one, whereas the reverse does not hold. The sudden emergence of a new behavioral mode (a new skill, grammatical structure, etc.) that characterizes a discontinuity does not imply that the old behavioral mode suddenly disappears. In fact, the two modes co-exist for a while (a situation which is clearly demonstrated by the fold in the cusp model). Eventually, the replacement of the old mode by the new may occur in a linear, continuous fashion. In that case, the frequency of the new mode increases gradually. It is likely that children will differ in the form of the transition between the old and the new mode, resulting in patterns that are clearly discontinuous in some children and considerably more continuous in others.

Another advantage of this approach is that the null-hypothesis can be formulated in various ways, dependent on the discontinuity criteria in question. Where some criteria may have continuity as the null-hypothesis, others may hold discontinuity as the starting point. Since the sufficiency question is no longer central in the analysis, this may lead to a much more subtle characterization of the developmental process under study.

4.3.3.2 (Dis)continuity as an inherently fuzzy concept

Although the idea of continuity and discontinuity as a continuum is radically different from the usual position in catastrophe detection studies, it is closely related to Fuzzy Logic. Fuzzy logic is a mathematical theory that can be used to program computers to “make decisions” based on imprecise data that can be positioned somewhere on a continuum. Fuzzy logic rests on the idea that all things admit of degree. Temperature, distance, length, friendliness, all things come on a sliding scale (McNeill & Freiberger, 1993). In contrast, traditional logic, set theory and philosophy use sharp distinctions. These sharp distinctions apply to the traditional continuity versus discontinuity discussion. A process is described as either discontinuous, or (if not all criteria are met) as continuous. Intermediate positions are traditionally not possible. We claim that allowing fuzzy classification in developmental psychology -more specifically in the discussion on continuity versus discontinuity- will lead to subtler and more informative analyses of developmental processes. Thus, in our view, continuity and discontinuity are distinct categories, but there is a range where they overlap and where the developmental phenomena have properties that are partly continuous, partly discontinuous.

A comparable point of view has been defended by Sternberg and Okaggaki (1989). In their article, they state that intellectual development is not either continuous or discontinuous, but that it is simultaneously continuous and discontinuous with respect to different dimensions of development. They propose that instead of asking the question “is development continuous or discontinuous?” we should ask “what are the sources of continuity and discontinuity in intellectual development?”.

4.3.3.3 The role of inter-individual differences

A general model of continuity and discontinuity must also account for inter-individual differences. In standard catastrophe theory, inter-individual differences are usually not analyzed. In principle, such
differences could be accounted for in terms of different values on the control parameters that define
the catastrophe at issue (two in the case of the cusp). Since catastrophe flags studies mostly focus on
the discontinuity region, the remaining differences almost automatically reduce to unsystematic
variation (noise). It should also be noted that not all catastrophe detection studies analyze the results
on the individual level, but perform their analysis on group averages. In these studies, inter-individual
differences are treated in the same way as intra-individual variability, namely as noise.
Assuming a dynamic systems approach, there are many different explanations to inter-individual
differences. Since the level of the developmental variable under investigation is a result of the dynamic
interaction between the developing child and the interaction with the environment, causes of individual
differences can be located in all control parameters that drive the process of development. However,
as we mentioned before, in order to account for discontinuities in some subjects, a general
(population) model of the developmental process has to incorporate discontinuity. Only if all subjects
show continuous trajectories, can a continuous population model be considered sufficient to describe
development. If the trajectories of some children show more characteristics of discontinuity, this can
be explained by the occurrence of a bifurcation, parallel to catastrophe theory. In the case of such a
bifurcation, there are two modalities (extreme positions) – a continuous modality that some subjects
display and a discontinuous one that can be observed in other subjects.
The study we have presented shows -in addition to several interesting similarities in the individual
trajectories- striking inter-individual differences. Not only does the timing of the first major increase
differ (for instance Berend has his first major increase very early on in his trajectory while Heleen
shows a relatively long first phase) but more interestingly, the shape of the transitions differs across
subjects. For instance, merely on the basis of visual inspection, the curve of Heleen shows a sudden
increase, while Berend’s highest levels increase more gradually. This is also reflected in the individual
results of the tested criteria. For instance, with the absolute distance criterion, Heleen showed low p-
values with the absolute distance criterion but much higher p-values with the peak criterion, while for
Lisa, the results were reversed. A similar finding is presented by Ruhland (1998). While some children
in his study show a sudden increase (e.g. subject Peter), others "take it easy and gradually increase
their production of function words" (pp. 85). In addition, Ruhland found differences between the
absolute numbers of instances of function words, for example Abel has an average of 150 tot 200
pronouns, while Daan produces between 150 and 400 pronouns per session. In our own preposition
data we have also seen that Lisa shows much higher levels in her first phase (thus before the first
major increase) than the other three subjects.
The finding of inter-individual differences in language development is hardly surprising (see Shore,
1995, for an overview). In fact, in the domain of language development inter-individual differences are
well documented. For instance, in the literature on phonological development, individual differences
are considered in the light of Universal Grammar (Beers, 1995). Also, the difference between two
learning styles (holistic versus analytic) is intensively studied (see for overview Bates, Dale and Thal,
1995). Because of these inter-individual differences, emphasis should -in our opinion- first be on the
individual level. Inter-individual differences also appear clearly in the form of the discontinuities in the
case study. Two of the children showed a possible, underlying discontinuity in the form of a sudden
peak of abundant preposition use (which is not artificial or ritualized) in an otherwise linear trajectory, whereas another child showed the discontinuity in the form of a clear step-wise pattern. A fourth child showed discontinuity only with the maximal performance criterion and the related criterion of maximal-minimal difference (note that the maximal level criterion is the only one for which all four children showed evidence of a discontinuity).

4.3.3.4 Towards a more general approach to (dis)continuity
The presentation of these empirical results should be seen in the light of a more general theoretical discussion we wish to open with regards to continuity and discontinuity. First of all, we argued for the incorporation of intra-individual variability in this discussion. Secondly, we have addressed two assumptions that play a major part in catastrophe detection studies. The first is that the distinction between continuity and discontinuity is usually conceived of as a forced choice between two contrasting, mutually exclusive positions. The second assumption is that the null-hypothesis in catastrophe detection studies is always formulated as a continuous process. As an alternative approach we have proposed to conceive of continuity and discontinuity as the extreme positions of a continuum. Instead of reporting whether a process is continuous or discontinuous, discontinuity is viewed as a collection of characteristics. These characteristics can be transformed into testable criteria that are compared with characteristics produced by specific null-hypotheses of (dis)continuity. With regard to inter-individual differences, we have argued that the individual level should be the starting point of the analysis and must precede eventual generalization over populations and processes. We are convinced that the incorporation of these elements leads to a new approach in the study of continuity and discontinuity that is able to explore the richness and the intriguing nuances of developing systems.