Chapter 7

Conclusion and discussion

7.1. Introduction: variability and ambiguity

This thesis is composed of five articles that touch on fundamental and methodological issues in developmental psychology. The themes of these articles are: intra-individual variability, ambiguity and discontinuity. Although these themes refer to distinct problems, they are strongly interrelated. This chapter focuses on discussing these interrelations and summarizing the results of the studies so far.

As all scientific disciplines, developmental psychology has several central assumptions. For instance, there is a strong focus on the gradual aspects of change, which is expressed in the fact that many developmental studies show average growth curves of a group of subjects. Underneath, there is the assumption that growth is in essence gradual. However, recent time serial data sets (sets of repeated measurements of the same individual through time), show a remarkable lack of smoothness. See for instance figure 7.1 (crying percentage, de Weerth, 1998). In this figure, crying behavior of one infant is reported in the percentage of time the child cried.

![Infant E](image)

Figure 7.1. Crying percentage as a percentage of the observed time (infant E), (taken from de Weerth, 1998)

The graph shows a distinct downward trend, crying behavior obviously decreases with time. However, the individual trajectory is characterized by strong fluctuations. While some days crying is virtually absent, other days show fairly high percentages of crying. A similar fluctuating patterns are shown in other developmental studies, such as function words (Ruhland, 1998), motor and mental development (Freedland & Berthenthal, 1994), sleeping and waking pattern in infants (Dittrichova, Paul,
Tautermannova & Vondracek, 1992) and visual behavior (Canfield, Wilken, Schmerl & Smith, 1991). As it seems, variability is an empirical reality that turns up each time a data set carries sufficient individual measurements through time.

A second central assumption in (developmental) psychology is that psychological categories are essentially explicit, in the sense that they are able to describe the observed behaviors in distinct true categories. Behaviors are assumed to belong to “crisp” categories that are mutually exclusive. For instance, a word is either a verb or not a verb. An infant has either acquired the concept of object permanence or not. A child either uses a certain mathematical strategy or not. However, in practice, observers are often confronted with difficulties when categorizing developing behavior that is essentially ambiguous. For instance, a child may use a word that has several characteristics of a certain category, but not all. An infant that is still in the one word stage may utter verbs as if they were nouns. These two assumptions of measuring developmental change are the focus of this dissertation.

Traditionally, the problems of intra-individual variability and ambiguity are being dealt with as problems of measurement. In the case of variable scores, psychologists take refuge in the concept of measurement error, while in the case of categorization problems the validity and reliability of the observational scales is being questioned. As I describe in the chapters 2 and 3, intra-individual variability is traditionally considered to be the result of measurement error. This error-hypothesis is a concept from true score theory (Lord & Novick, 1968, Cronbach, 1960, Nunnally, 1970) and is deeply rooted in psychology. True score theory is based on the assumption that every psychological measurement is subjected to measurement error. The “true score” is conceived of as an abstract entity that represents the factual value of a variable. It is assumed that in practice, testing procedures are never perfectly consistent. Theories of test reliability have been developed to estimate the effect of inconsistency on the accuracy of psychological measurement. The basic starting point is that all test scores reflect two sorts of factors: (1) factors that contribute to consistency (stable characteristics of the individual or the attribute one is trying to measure), and (2) factors that contribute to inconsistency (features of the individual or the situations that affect the test scores but which have nothing to do with the attribute being measured) (see Murphy & Davidshofer, 1991, pp. 75). Note that with this definition, all instability is conceptualized as being error. This conceptual distinction can be represented in the following equation:

\[
\text{Observed test score} = \text{true score} + \text{measurement error}
\]

or:
\[X = T + e\]

Measurement error is assumed to be unrelated to the true score, which leads to the conclusion that the mean of errors equals 0. Since these errors are independent of the true score, they are also assumed to be symmetrically distributed around the true level. Thus, in practice fluctuations as found in time serial data are considered to be added to development from an external source rather than
Conclusion and discussion

being a characteristic of the process of change itself. By averaging over the fluctuations, the true underlying level can be approached. The chapters 2 and 3 of this dissertation have discussed this problem of intra-individual variability and true score theory and have concluded that the strong focus on gradual aspects of change and stability across development is no longer justified in the light of recent theoretical insights and empirical findings. In these chapters, I have focused on a radically different perspective that is posed by dynamic systems theory (Thelen & Smith, 1994; van Geert, 1994).

The starting point of dynamic systems theory is that a developing system is maintained by a flux of energy. Every developing system is constrained by limited resources, such as memory, attention motivation, etcetera. This system is in constant complex interaction with its environment and internal sources. Its multiple interacting components produce one or many self-organized equilibrium points, whose form and stability depends on the systems constraints. These equilibria are called attractors. Growth is conceived of as an iterative process, which means that the present level of development depends critically on the previous level of development (van Geert, 1994). Furthermore, growth can be expressed as a mathematical equation (the logistic growth equation) in which growth is dependent on growth rate $r$, a power parameter $p$, the carrying capacity $K$, and feedback delay. Also models of connected growers can be formulated, expressing competitive or supportive relationships.

Dynamic systems theory aims at explaining two levels of development with the same principles (see Thelen and Smith, 1994). The first level is the view from above, or the “grand sweep” of development. At this perspective, we see global structure, similarities across subjects. For instance, when infants learn how to walk, they perform roughly the same behavior. The second level of development takes a view from below. From here, we observe the messy details: behavior that is variable, fluid, and highly context dependent. We see, for instance, that not all infants use the same strategies when learning how to walk, and show large variability across time. This variability is viewed as an important internal developmental characteristic and not of something externally added to the process of development (such as noise). The theory radically rejects the automatic retreat to the error hypothesis and claims that variability bears important information about the nature of the developmental process. Dynamic systems theory stresses the importance of the context in which the behavior is displayed. Development takes place in real-time and is considered highly context-dependent. Therefore, it can be compared with an evolutionary process, which is also mindless and opportunistic. Thelen and Smith state: “There is no design, no blueprint, no pre-given specifications for the species that emerge. There is no end state, only context specific adaptations” (pp. 144). They also agree with the classical Darwinian emphasis on variability as the source of new forms. They state: “We believe that in development, as in evolution, change consists of successive make-do solutions that work, given abilities, goals and history of the organism at the time” (pp.144). Variability is considered to be the result of the systems’ flexibility and adaptability to the environment. From a dynamic systems angle, variability has been viewed as both the source of development (for instance Berthenthal, 1999; Thelen and Smith, 1994) and an indicator of a specific moment in the developmental process, namely in the presence of a developmental transition (van der Maas & Hopkins, 1998).
Traditionally, in case of problems in categorizing developing behavior, the concept of interobserver reliability is introduced. This concept is based on the risk of observer bias, which is the tendency of the observer to confirm one’s hypothesis by reading too much into naturalistic observations. In other words, the aim is that the observation is as objective as possible, and not influenced by individual observer interpretation (see general handbooks such as Shaffer, 1989). Thus, differences between observers are conceived of as one specific kind of error. Note that the concept of interobserver reliability is also closely related to the error hypothesis. By comparing scores (or transcripts) of two independent observers, this error can be estimated. If agreement between observers is too low (for instance below 0.75, see Murphy and Davidshover, 1988, pp. 103) the procedure of the study is assumed to be unreliable and has to be adjusted. However, in practice, reliability is critically dependent on the concept of true behavioral categories that are mutually exclusive. This practice is also criticized from the dynamic systems angle. For instance, in a criticism of Wimmers (1998), Thelen (1998) questions the method that he used to determine whether infants reached without grasp or with grasp. According to Thelen, these categories impose arbitrary criteria on what is in fact continuous behavior. She suggests that here may have been intermediate forms (regions of inaccessibility in the cusp model) but we would never know it, because by definition the system must live in one mode or another (pp. 154).

The concept of ambiguous categories is closely related to Fuzzy Logic (McNeill & Freiberger, 1993). Fuzzy logic rests on the idea that all things admit of degree. Temperature, distance, length, friendliness, all things come on a sliding scale. In contrast, most of traditional logic, set theory and philosophy have prescribed sharp distinctions. This sliding scale yields greater precision and can quantify degrees of truth. In modern fuzzy logic, “objects” (observations, objects, properties, etc.) are always assigned a Degree-of-Membership (see Ross, 1995, for a particularly clear technical introduction; see further Kosko 1993, 1997; Nguyen & Walker, 1997; von Altrock, 1995; McNeill and Freiberger, 1993, provide a highly accessible introduction). In mutually exhaustive classes, objects have a degree-of-membership of either 1 or 0 (for instance, a word is either a preposition or not a preposition). In fuzzy logic, an object can have any degree of membership between 0 and 1. Maximal ambiguity arises if an object has a degree of membership equal to 0.5. In this case the object has a position at the exact mid-point between the two extremes. When this is the case, its ambiguity is equal to 1. If the degree of membership is either 0 or 1, ambiguity is 0.

7.2 Aim of this thesis

If both variability and ambiguity are important characteristics of a developing system, there are two lacunas in developmental psychology. The first lacuna is that there is relatively few empirical knowledge on variability (fluctuations) and ambiguity. The only knowledge about these phenomena is in the context of measurement error and reliability. However, in this context variability and ambiguity are removed from the study, in the sense that they are being “averaged out”. The second lacuna that might be responsible for this lack if knowledge is that there is a lack of technical tools, in the form of
Conclusion and discussion

Statistical methods and procedures, which can be used to describe and analyze patterns of fluctuation and ambiguity. It is the main goal of this thesis to contribute to a solution to both of these lacunas. Since empirical knowledge critically depends on the method of study, I have first contributed to the methodology of investigating these two phenomena in a developing system, by proposing various simple but powerful techniques. Secondly, I have started a study of variability and ambiguity in one specific domain of development, namely early language development.

7.3 Methods to study variability in time serial data

With regards to our first goal, I have aimed at the development of various simple techniques that can be used to analyze patterns of intra-individual variability and ambiguity. In spite of the potential importance of this variability phenomenon (as dynamics systems theory claims), the standard methodological toolkit of the developmental psychologist offers hardly any instruments to study it in a structured fashion. In this dissertation, I have discussed several well-known techniques and added several new procedures for this purpose. The details of the techniques are described in chapter 3.

In chapter 2, I have demonstrated the use of the boxplot (see figure 7.2). The boxplot is a well-known graphic representation that displays the spread of scores within a defined time frame. It provides information about the median and the range of a set of observed scores. In a boxplot, 50 percent of all observed values are depicted as a box. The upper boundary of this box depicts the end of the third quartile, while the lower boundary of the box shows the end of the first quartile. The whiskers above and below the box reach to 1.5 times the boxlength (in observed values). Values that exceed this boundary are called outliers (till 3 times the boxlength) or extreme values (over 3 times the boxlength). This technique is useful for visualizing variability over a short period of time, where the role of
developmental increase is judged to be small. In data that cover a larger portion of time in which development is assumed to influence the distribution through time, the boxplot is not suited. For these instances (actually for all instances in which one is interested in how variability changes over time) I show alternatives in chapter 3: the moving min-max graph and the progmax-regmin graph. The underlying principle is the use of moving windows: a time frame that moves up one position each time. Each window largely overlaps the preceding windows, using all the same measurement occasions minus the first and plus the next. For instance, for every set of seven consecutive measurements we calculate the maximum and the minimum values. This is done by way of a predetermined moving window, such that we obtain the following series:

\[ \text{max}(t1..t7), \text{max}(t2..t8), \text{max}(t3..t9), \text{etc} \]
\[ \text{min}(t1..t7), \text{min}(t2..t8), \text{min}(t3..t9), \text{etc}. \]

Figure 7.3.AB. Application of the ordinary min-max method (top) and the Progmax-regmin method (bottom) to Heleen’s spatial prepositions.
Instead of displaying measurement points as simple dots, the moving min-max graph presents a score range for each measurement occasion. Thus, instead of a single line graph, the data are presented in a bandwidth of scores (see figure 7.3A). We can also plot some form of central score, or even “altitude lines” to indicate the position of the other scores in the predefined window.

The progmax-regmin graph is a variant of the moving min-max graph is (see figure 7.3B). This type of representation also displays a score range or bandwidth, but specifies the range of the entire time period of the study, where the moving min-max graph analyzes local windows. For instance for a dataset that consists of 20 measurements, the window in the case of the progmax-regmin graph is determined as:

\[
\text{max}(t1-t5), \text{max}(t1-t6), \text{max}(t1-t7), \text{max}(t1-t8), \text{etc} \\
\text{min}(t1-t20), \text{min}(t1-t19), \text{min}(t1-t18), \text{min}(t1-t17), \text{etc}.
\]

In addition, we demonstrate a technique that is able to detect sudden increases of variability: the critical frequency method. This method is developed in the field of motor coordination (e.g. Verheul & Geuze, 1999) and can be applied to developmental research in general. The method is based on the following assumptions. First of all, a system is supposed to be relatively stable over some initial period of time. Second, this period must be followed by a period in which this system becomes “unstable”, which results in large variability. The aim of this method is to establish if, and at what exact moment, the system loses its stability (see figure 7.4).

\[\text{Figure 7.4. Application of the critical moment method to Heleen’s preposition data.}\]

In the application of this technique, we have to bear in mind again that we are dealing with a variable that shows a considerable increase, i.e. a growth trend. Since we do not want our general trend to influence the variability measure we first have to detrend the data. Consequently, we calculated the residuals of the original data based on a flexible regression model. The critical period, was determined by calculating a moving standard deviation (using a moving window of 5 observation points) on these
residuals. It is now possible to establish at what moment the variability in the system “critically” increases. This is the case when the moving SD exceeds the critical value (which is the upper limit of the reliability interval) for at least 6 consecutive moving SDs. I applied this technique to the data of Heleen’s spatial prepositions, and plotted both the preposition data and the moving SDs on the residuals of these data in figure 7.4. As can be seen, the critical value is exceeded once at measurement point 29 (September 16th) and exceeded again at measurement point 37 (October 29th). But only after measurement point 42 (December 1st) is this value exceeded for the seventh consecutive time, which means that showing a significant increase in variability occurs after measurement 42 (December 1st). Also, a technique is proposed that is based on the symmetric distribution assumption of the measurement error hypothesis: the moving skewness graph. This methods rests on the assumption that measurement error is distributed symmetrically around the true score. When studying the distribution characteristics, we are especially interested in the skewness of the distribution since the skewness can give us information about the degree of consolidation of the acquired developmental variable. For instance, when a child begins to discover that he or she can use prepositions to express spatial relations, we expect outliers on the positive side of the distribution. It is highly likely that the child uses this new linguistic category (in this case spatial prepositions) in outbursts, but most of the time this new category is not used and the child simply sticks to his or her usual repertoire (for instance by pointing to a location).

Figure 7.5. Negative skewness, also called left skewness (top) and positive skewness, also called right skewness (bottom).

It is important to note that the study of variability also implies some methodological considerations, which I will only mention briefly. First, the techniques are especially useful for visualizing and describing intra-individual variability in time serial data on child development (minimally 10 to 15 measurement points). Studying variability requires time serial data, which must first be studied at the individual level. Also, measurement points should be placed close together, dependent on the variable that is studied. Variables that are known to develop rapidly (such as language development) should
have sufficient measurement points to detect short term variability in the vicinity of possible developmental jumps. In fact, a case study (Heleen) showed that the addition of six so-called “intensive periods” contributed significantly to the pattern of variability that resulted from the study (see chapter 3). Therefore, the measurement design should be scheduled in such a way that both long(er) and short term variability can be detected.

Again, in order to eliminate the influence of the general increase, it is important to detrend the data, using a trendline, for instance by calculating the residuals for a regression line. Subsequently, the distribution of these residuals can be studied for skewness, computing the skewness factor on moving windows, which move up one position at a time. The result is a flexible symmetry analysis for each part of the developmental trajectory (see figure 7.6).

![Figure 7.6. Application of the moving skewness graph to Heleen’s spatial prepositions](image)

### 7.4 Resampling techniques

In this thesis I have repeatedly demonstrated the use of resampling techniques, for various purposes (for a general discussion of the bootstrap and resampling methods, see for example Good, 1999 and Manly, 1997). For instance, I have used these techniques to test the “significant contribution” of the intensive day-to-day measurement in Heelen’s data set (in chapter 3) and the “significance” of the empirically found overlap-ratio of two independent observers (in chapter 6). I have also demonstrated a resampling procedure that aims at detecting “significant” characteristics of discontinuity (chapter 4). Note that I have used the term significant in quotation marks. This term is used if the resampling procedure has shown that there is a very low probability (below 5%) that the data are produced by a chance model simulated in the resampling procedure. In fact, resampling procedures can be used to test an infinite variety of research questions.
Chapter 7

The procedure itself is extremely flexible and is assumption free. In the case of analyzing time serial data, this is an important feature, since the standard statistical methods (such as multilevel analysis) are not suited for data in which the lowest level measures are not independent. However, it should be noted that the testing model that is built in order to employ a resampling procedure is essential. In fact, a resampling procedure is as good as its model. If a model is formulated “too weakly”, its results are meaningless.

The resampling procedure with regards to continuity versus discontinuity deserves some special attention. In dynamic systems theory, 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, 1997). In spite of its importance, the phenomenon of intra-individual variability is an underexposed aspect in the discussion on continuity versus discontinuity, mainly since there are no methods to study these phenomena in an integrated approach. However, (dis)continuity and variability are inherently intertwined. In chapter 4 we have proposed to describe discontinuity as follows: discontinuity 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. In this chapter I have demonstrated a procedure that tests for the existence of two variability-centered criteria. The first is the existence of two clearly distinct sub-phases or sub-patterns in a trajectory. 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. 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 resampling procedure consists of two steps. The first step consists of an analysis of the individual trajectories, based on bootstrap and resampling procedures (the procedures were carried out in Microsoft Excel, by means of a statistical add-in, Poptools, Hood, 2001). The question I addressed is: do the individual trajectories show discontinuities in the developing range? Second, I 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. The question I 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 was estimated on the basis of the observed data. The continuous model was 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.

7.5 Methods to study ambiguity

In the second part of this thesis (chapter 5 and 6), I have also demonstrated the use of several methods to analyze ambiguity and the effect that it has on our results. In chapter 5, I have described
Conclusion and discussion

how these choices made with the transcription of unintelligible speech may lead to an underestimation of indices of utterance length. A procedure is suggested to investigate the magnitude of this underestimation by using the original values (in which xxx and & are excluded and which underestimates utterance length) in combination with a worst-case-scenario. In this worse case scenario a fixed value procedure is used in which the utterance length is explicitly overestimated by assigning each instance of xxx and & with the value of 1. This way it is possible to acquire a bandwidth (between the original values and the values that result from the fixed values procedure), between which the values vary dependent on transcription practices. This also gives an impression of the maximal influence the partially interpretable utterances have on the developmental trajectory.

In chapter 6, I have demonstrated how a lack of inter-observer reliability may be caused by an intrinsic ambiguity of the developing language system. In this chapter, it was analyzed how the original overlap can be transformed into a measure of ambiguity based on fuzzy logic principles. The suggestion to treat linguistic categories as fuzzy sets is quite uncommon in the language acquisition literature, to say the least. However, I have suggested that this theoretical application is not far-fetched. For instance, the research dealing with “filler syllables” (also called Prefixed Additional Elements, PAE) encounters a similar issue when discussing the status of this phenomenon (see for instance Veneziano and Sinclair, 2000). The question is whether these filler syllables can be considered grammatical elements or not. One group of authors (e.g. Dolitsky, 1983; Peters, 1990; Veneziano, Sinclair & Berthoud, 1990; Scarpa, 1993; Simmonsen, 1993; Kilani-Schoch & Dressler, 2000) linked filler syllables more specifically to the child’s development of grammatical morphemes, considering them “an intermediate form on the way to grammatical morphemes” (Veneziano and Sinclair, 2000; pp. 463). Thus, although the filler syllables are not considered to be articles, prepositions or auxiliaries yet, they take an intermediary position. In fuzzy logic terms, they might acquire the value of 0.5 (or any other value between 0 and 1) on the dimension “grammaticality”, where the value of 0 stands for “not-grammatical” and 1 for “grammatical”. In chapter 6, we have illustrated how these fuzzy logic principles can be used to estimate the levels of ambiguity based on overlap between independent observers.

7.6 Initial findings in early language development

With regards to the second goal, namely to contribute to initial knowledge on variability in early child development, I have conducted a study in the domain of early language development. The domain of language development is especially suited for the study of variability for several reasons. First of all, because the study of variability requires a relatively large collection of measurement points per individual, the measurement procedure itself must be as non-obtrusive as possible. The collection of spontaneous speech samples easily meets this demand. Secondly, language is known to develop relatively quickly and shows a rapid increase in its complexity. Thus, meaningful data sets can be collected in relatively short periods of time (about one to two years, on average). Also, the domain of language development is suited for such a quantitative approach, since it poses the categorization
issue in a natural way (for instance “is this word a preposition?”). Although language development is predominantly studied from a qualitative, structural angle, language easily provides quantitative data, which can be plotted and used for further calculation.

In this pilot, I have studied intra-individual variability in quantitative data of four infants (Heleen, Jessica, Berend en Lisa). Data sets of these infants consisted of 45-55 repeated observation (spontaneous speech samples) within the course of a year (from roughly age 1;6 to 2;6). The details of the study are reported in the chapters 3 and 4.

![Graph of Jessica's MLU-w](image)

![Graph of Heleen's MLU-w](image)

Figure 7.7A. Individual trajectories of MLU-w of Jessica (top) and Heleen (bottom)

In chapter 2, I describe a pilot study in which intra-individual variability in early language development is studied on a very short time scale. This study was performed on spontaneous speech samples from two young children (Eva and Jan, age 2;4 and 2;8). Short-term variability was studied in terms of Mean Length of Utterance (MLU-m, Brown, 1973), a widely used language measure. The data of both
children showed considerable variability in MLU, even between measurement days. The two children in the case-study even fall into several MLU-stages within a period of weeks. Although this finding may be surprising at first sight (especially with regards to the assumed meaning of MLU in the literature) it corresponds to the finding of de Weerth (1998) in a different developmental domain. In this study, infant behavior (negative and positive emotionality and physical contact) was found to be highly variable both between weeks as on a day-to-day basis. In fact, there is no primal reason to suspect that behavior that is variable at the semi-short interval is not also highly variable at a shorter time scale.

Furthermore, the longitudinal study showed that intra-individual variability was high for all four subjects (prepositions and MLU-w) and at all time scales (see figures 7.7A and B). In the case of Heleen and Jessica, this variability pattern has some further interesting characteristics. First, visual inspection of these two distinct language variables shows that each has a specific variability pattern: while MLU develops slowly, with an increasing variability, spatial prepositions show a jump-wise pattern (see figures 7.7A and B).

Additionally, in the case of Heleen, the distribution of fluctuations was tested by means of a moving skewness graph. As it turned out, the degree of symmetry is not equal in each part of the developmental trajectory. Instead, an oscillating pattern is observable, with four distinctive parts. The data start out with a near symmetric (only slightly positively skewed) distribution, which turns into a strongly negatively skewed distribution from July to midway September. Then, from midway
Chapter 7

September to the end of November, a fairly strong positive skewness is detected, again followed by a part with only moderate negative skewness factor, after November. Note again that, in the standard approach, error is supposed to be symmetrically distributed around the central tendency, which is supposed to represent the best estimation of the true underlying variable. The oscillating skewness patterns that are found in Heleen’s spatial prepositions do not support this symmetry-assumption. When examining this pattern even closer, it turns out that there is a positive skewness right before the moment of the jump, which turns into a negative skewness at the moment of the jump itself. (see figure 7.6) Although this result is speculative, it suggest that variability is not symmetric but related to a specific moment in the trajectory (the jump).

In addition, I have studied variability in language development in relation to the question of whether there are discontinuities in the development of spatial prepositions. First, the formal analyses have confirmed that the time at which the distance between the two distinct phases is maximal occurs around point zero, which was defined as the moment of the first major increase in the data, based on visual inspection. If the distance between the two sub-phases was estimated on the basis of the functional or average level of preposition production (which is what the linear estimation procedure accomplishes) there was no support for the hypothesis that the data reflected a discontinuity (see table 7.1). However, if this distance was estimated on the basis of the local maximum (which indicates the optimal performance at that point in time), there was strong support for the discontinuity in all subjects. With regard to the second variability-centered criterion, I found that two subjects showed an unexpectedly large peak. The meta-analysis that combines the results of the subjects showed a significant result which indicates that the four subjects combined show indications of discontinuity.

Table 7.1

<table>
<thead>
<tr>
<th></th>
<th>Distance (on funct. level)</th>
<th>Distance (on local max)</th>
<th>peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berend</td>
<td>0.40</td>
<td>0.03</td>
<td>0.42</td>
</tr>
<tr>
<td>Heleen</td>
<td>0.08</td>
<td>0.01</td>
<td>0.38</td>
</tr>
<tr>
<td>Jessica</td>
<td>0.25</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>Lisa</td>
<td>0.54</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>Meta analysis</td>
<td>0.19</td>
<td>_</td>
<td>0.01</td>
</tr>
</tbody>
</table>

These analyses have demonstrated that variability and discontinuity are interrelated at the moment of the first major outburst (peak, jump) in the data. However, there are obviously different types of discontinuity, dependent of the combination of criteria for each subject. For instance, in the case of
Heleen almost all distance criteria were significant, while the peak criterion was not. In the case of Jessica, the peak criterion was highly significant while only part of the distance criteria were significant. This means that Heleen's data can very well be characterized by a two stage model, while in the case of Jessica, the unexpected peak is a stronger criterion and might be responsible for the other types of discontinuity as found in the data.

In this thesis it is argued that intra-individual variability is an intrinsic characteristic of a developing system. A system that has reached its "end-state" (for instance the end-state of language development) should not display this kind of variability to this degree. On the other hand, we must consider the possibility that the end-state of development is also not "stable", but a state that is "dynamically stable" in the sense that the produced language still shows a considerably variable range dependent on situational factors. This range (of variability) should, however, be smaller than the range of preposition use in young children who are still in the process of development.

There are indications that this is indeed the case for at least one aspect of early language development: children appear to be considerably more variable in their use of spatial prepositions than adults. A resampling procedure showed that, on average, variability of the four infants in the case study (Heleen, Jessica, Berend and Lisa) were larger than that of two adults samples. These results indicate that it is highly likely that the adult use of prepositions-in-context shows a lower variability than that of the children.

With regards to ambiguity, I report results from a case-study (Heleen) showing that the curves of 1-, 2- and 3-word utterances and MLU reflect an underlying developmental pattern and are not just the result of increased interpretability. I have defended that ambiguity can be quantified in terms of uninterpretability and have provided information about the process of development. Also, a case study on the classification of this early child speech (utterance by utterance) into two mutually exclusive categories “preposition present” or “no preposition present” showed an inter-observer reliability (overlap ratio) of 0.87. Following a worse-case scenario in which all the disagreement is considered to result from ambiguity, this value of 0.87 van be transformed into a level-of-ambiguity of 0.72, which is fairly large (0 means no ambiguity, 1 means maximal ambiguity).

### 7.7 Recommendations

The study of variability and ambiguity in early child language is still in its infancy. In fact, the field is open to a wide variety of research questions. However, there are two specific recommendations for further research, that follow from this thesis.

The first recommendation deals with the way a quantitative method such as presented in this thesis is compatible with a qualitative approach of linguistic theories. Traditionally, linguistic theories have been inspired by the fact that language is highly organized. Language has a structure in two senses: surface structure (sentence) and deep structure (underlying structure in the shape of a syntactic
tree). The idea of deep structure is a crucial element of generative grammars. The question is: how does a child acquire the rules that underlie this structure? This “learnability problem” has been the focus of a large portion of linguistic research (Pinker, 1988; Berwick, 1987; Clahsen, 1992 see Gillis and Schaerlaekens, 2000). The two extreme answers to this problem have been formulated by learning theories and nativist theories. Although the first set of learning theories such as Skinner’s are no longer considered as an adequate explanation of language development, it might be questioned whether it is justifiable to discard the influence of learning mechanisms on language acquisition altogether (van Geert in Gillis and Schaerlaekens, 2000). On the other extreme, Chomsky and his followers have argued that the acquisition of language is impossible without innate linguistic knowledge. This body of knowledge, known as Universal Grammar is a set of principles (constraints on the architecture of grammars) that sets constraints to the range of all possible human languages. The answer to the learnability problem is that languages do not differ in terms of their core grammatical architecture. In this core architecture, the universal principles are parameterized in such a fashion that the designs of specific languages vary over a limited range, each of which implements the principles differently.

In this thesis, I have not dealt with the theoretical discussion between learning theories, nativists and other theoretical viewpoints. The point I want to make here is that structural accounts of language development do not need to be conflicting and can, in fact, be integrated with our quantitative approach. We agree with Thelen and Smith (pp. 26) that Chomsky’s theory is a strictly formal account of linguistic structure, excluding meaning and saying nothing about real language behavior and its development. The fact of development is not explained by a list of innate ideas. Or as Thelen and Smith state: “All the facts possible about the architecture of the end-state won’t tell you how it got there” (pp 36).

In order to illustrate how this quantitative approach can be related to a qualitative question, I have conducted a pilot study on the acquisition of two types of early prepositional phrases (van Dijk & Krikhaar, 2002 unpublished raw data). In this study, I describe the developmental trajectory of two forms of prepositional phrases. These are PE(alone) (this means that the child only utters a prepositional element and no other lexical elements) and PE-NP (in which the prepositional element is used in combination with a nominal phrase which is the subject of the Prepositional Element - the adult form). In this study, the developmental trajectories of these different types of prepositional utterances of four subjects (Heleen and Lisa (see this thesis) and Matthijs and Peter (Groningen Dutch Corpus, Krikhaar & Wijnen)) were plotted. The trajectories of these four subjects showed interesting characteristics. First, all subjects had a stage in which many different types of prepositional utterances were used simultaneously. Secondly, while two subjects (Lisa and Matthijs) showed a gradual growth of the adult (PE-NP) form, two other subjects showed a jumpwise pattern (see figure 7.8).

Also, intra-individual variability was large in all subjects. Furthermore, there were indications that there is a competitive relation between the PE-alone form and the PE-NP construction. This pilot study, we concluded that the development from PE-alone to the PE-NP construction seems to be the result from
a complex interaction between different elements (competitive and supportive relations with other types prepositional utterances, MLU, productivity, other related categories such as Verb-Particles). Further research has to be conducted in order to specify these relationships. However, this illustrates how a dynamic systems model leads to the analysis of different types of relations between linguistic categories than the usual generative approach (MacWhinney, 1998).

Figure 7.8. Examples of the PE-NP construction in preposition use (proportional, 1 is 100% of all prepositional utterances). Subjects Lisa and Matthijs show gradual growth patterns (top), and Heleen and Peter with jumpwise patterns (bottom).

With regard to ambiguity, I have formulated two further predictions in chapter 6 that have yet to be tested empirically. The first prediction concerns the use of “intermediate” categories in a scoring system onto inter-observer reliability. Instead of classifying child utterances in one of two extremes
("preposition-present" and "no-preposition-present), the utterances can take an intermediate position 
("preposition-possible"). In the example, these utterances neither belong to the category preposition-
present nor to the category preposition-not-present but belong to the category preposition-possible or 
ambiguous-preposition. We might expect that using such an intermediate category improves the 
overlap percentages to some degree since the ambiguity is acknowledged in the coding system. One 

further advantage of such an approach is that results can be reported dependent on these degrees of 
certainty. For instance, a graph can be plotted with a single line that reports all prepositions (or any 
other linguistic category) in which the observers are certain and a line that incorporates all "possible" 
instances. Comparing both lines expresses to what degree the results are sensitive to the uncertainty 
of the observers (see also van Dijk and van Geert, submitted -chapter 5 of this thesis-, for a 
generalization of this line of reasoning). On the other hand, although it would certainly be informative 
to ask raters how certain they are about their rating, this procedure might eventually only replace the 
problem. Instead of having an assumption of two mutually exclusive ("crisp") categories, i.e. 
preposition and not-preposition, we now have the assumption of three mutually exclusive categories, 
namely preposition, ambiguous and not-preposition. If the categories are truly ambiguous or fuzzy, 
inter-observer (dis)agreement will now pertain to three categories instead of two. It is unsure whether 
this procedure results in higher reliability values (since the scoring procedure fits better with the 
ambiguous nature of the data) or if disagreement increases between observers with regards to the 
application of this third category. This question can only be answered in an empirical study in which 
these different scoring procedures are being compared.

The second recommendation with regard to further research on ambiguity concerns the expectation 
that in the course of development inter-observer reliability increases since ambiguity decreases, which 
I mentioned previously. The subjects in the case study were between one-and-a-half and two-and-a-
half years old during the time of the observations. We expect a higher agreement between observers 
for the observations that were made later in the trajectory. In addition to the improvement of the 
phonological system, older children have better defined and stable linguistic categories, which result in 
a decreased intrinsic ambiguity. This prediction can be tested in a developmental study on how the 
level of ambiguity changes over time.

In this thesis, I have elaborated on the concepts of variability and ambiguity. I have argued that these 
two concepts are important, not only in the study of early language development, but also in other 
domains of child development. In this thesis, I have initiated a first step toward the development of 
approaches and techniques that are suited to study these phenomena in developmental data. This 
implicates a different view on development, which embraces ambiguity and variability as intrinsic 
characteristics of a developing system. Hopefully, it will be used as a source of inspiration to the 
development of new hypotheses and future research.