From Language Learners to Dynamic Meaning Makers: A Longitudinal Investigation of Malaysian Secondary School Students' Development of English from Text and Corpus Perspectives

Chau Meng Huat

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

Department of English Language and Applied Linguistics
School of English, Drama, and American and Canadian Studies
College of Arts and Law
The University of Birmingham
February 2015

Abstract

This thesis considers how language development takes place over time by a group of 124 secondary school students of English. A series of five studies were conducted for this purpose using the tools and methods from corpus linguistics and written discourse analysis. Specifically, the thesis presents a detailed analysis of (1) how a set of function words (*that*, *to* and *of*) were used by these students over a 24-month period, and (2) how narrating practices concerning the structure of selected individual texts changed over time. The two distinct strands of investigation, both of which based on an inductive methodology, highlight, on the one hand, the extent to which there are common as well as unique aspects of language use observed across time and space (Francis et al., 1996, 1998) and, on the other, the role of human agency and meaning making practices in using linguistic resources over time and in shaping and constructing texts within and across individuals. Taken together, the overall inductive methodology and an emphasis on treating all instances of the conventionally labelled 'learner language' as equally valid features of natural human language use, show clear advantages over alternative approaches based on a deficit model.

For

my mentors – heroine and heroes in my academic life who took me under their wings

Susan Hunston, Khairi Izwan Abdullah, Nicholas Groom and the late Richard Pemberton

my family and most supportive friends

Ah Loi, Tean See Moi, Tow Teck Soon, all my eleven siblings with their spouses and beautiful children (plus my grand niece and nephew!), Hassan and Fara, Chye's family, Kamariah Ahmad, Lee Kok Yueh (especially), Su Hang (especially), Teoh Mei Lin (especially), Thang Siew Ming (especially), Jane Evison (especially), Dariusz and Klaudia Świerad (especially), Heng Chen Chuan, Rose Easaw Thomas, Jan and Frank Hardman**,** Stefanie Pillai, Susie Li, Arezoo and Masoud, and those friends who have been praying hard for my leading a blessed life

all my students, especially George Wong, *and teachers*

and

all the people I have met or read or heard about, and the incidents and things I have encountered in my life, including

this thesis

who has taught me not only about human language use but more significantly,

about what it means to be human

Acknowledgements

I would like to thank the Ministry of Education for the scholarship which supports three years of my PhD research at the University of Birmingham.

In addition, I thank the University of Malaya for the award of an SLAB/SLAI fellowship as well as the award of a Bright Sparks Fellowship in 2011.

My heartfelt thanks also go to Professor Michael McCarthy and Dr Suganthi John for agreeing to examine this thesis. I had an intellectually stimulating time during the *viva voce* examination.

I am fortunate to have two supervisors supporting my research journey. I owe a huge debt of gratitude to Dr Nicholas Groom. Nick provided helpful comments and constructive criticism for me on the early drafts of different chapters of this thesis.

I am most grateful to Professor Susan Hunston for sharing my excitement in this research and supporting me throughout my study. It is no exaggeration to say that Susan's supervisions were the most exciting and stimulating academic experience I have ever had. I consider myself extremely privileged, and immensely blessed, to have had the benefit of her guidance and commitment to my research. She was the first person on my mind when I was thinking of having a dedication page, which turns out to be an unusually long dedication note(!).

As the cliché goes, responsibility for any inadequacies which remain in the thesis rests firmly at the door of the author.

Table of Contents

Part IV

List of Tables

List of Figures

Chapter 1

Introduction

1.1 General aim of the thesis

The aim of this thesis is to make sense of how second language development takes place over time by examining the developing language of an under-researched population: young, secondary school students of English as an additional language. Relatively little has been published about these students to date, especially from a longitudinal perspective, making this a significant area of enquiry. Empirical data in the form of narrative texts written by these students at different points in time are considered and explored using corpus and discourse or textual approaches. In researching second language development, this thesis gives priority to evidence generated from the study of the process of internal change observed in the developing language, rather than through comparisons involving an external frame of reference or source of data.

This thesis, in other words, challenges the current, dominant practice in the study of 'learner language' which focuses primarily on the comparison between the language performance by the 'learner', reconsidered here as the *developing language user* or meaning maker (see **Section 1.2**), and that of the 'native speaker', or an idealized language competence, or an external point of reference. Taking the developing language as a dynamic meaning-making resource – a natural form of human language in its own right, the thesis presents ways in which changes in the developing language over time are studied without treating it as a deficit model against an external norm and without making an assumption about what is inaccurate. The extent to which this is a feasible and worthwhile attempt is duly assessed.

1.2 Key terms and concepts in the thesis

In this thesis, distinctions are made between (1) a 'learner' and a developing language user, and between (2) conventional language use and innovative language use.

1.2.1 '*Learner*' versus *developing language user*

I use the term *learner* (without quotation marks) in its general sense: 'we are all learners in some areas of our language and beginners in many others' (Hoey, 2005, p. 184). What has been a dominant practice in contemporary Applied Linguistics is, however, that those learning or studying an additional language after their first language(s), defined in this thesis as a second language, are labelled 'learners'. Thus, we have specific groups of people being reported in research as 'learners of English', 'learners of Mandarin', etc. This group of human population is often contrasted with another group of human population speaking the language, known as 'native speakers'. Thus, we talk about 'native speakers of English', 'native speakers of Mandarin', etc, and how language use by the failing 'learners' differs markedly from that of their 'native' counterparts.

Such a contrast is misleading, not least because the linguistic differences between the two are largely a matter of degree rather than absolute distinctions. The comparison is scientifically untenable as we are comparing the language or meaning-making resource of a bilingual or multilingual speaker through a monolingual, 'native-speaker' lens. Further, there are issues of individual identity and aspirations not adequately addressed when we draw a divide between a 'learner' of a language and a 'native speaker' of the language, assuming that the ultimate goal of the former is to be the latter. To avoid the stigma attached to the notion of the 'learner' as being merely a 'nonnative' or 'defective communicator' (Firth & Wagner, 1997, p. 285), this thesis considers and addresses this group of individuals as *developing language users*, including the secondary school students in the research to be reported herein. I also, interchangeably, refer to these young language users as emergent language users (hereafter ELUs). In this thesis, language that is usually referred to as 'learner language' is referred to as developing language or emergent (user) language.

It must be noted, however, that this is not merely an ideologically motivated deliberation. The reconsideration of the notion of 'learner' as a developing language user acknowledges two important facts about the user's engagement with language, based on the findings of the empirical studies to be reported in Chapters 4-8 of this thesis. First, these users, like other language users, exploit language to make meaning: they are evidently producing and using the language to communicate their intentions and thoughts. Second, they change their language use over time, just as any other language user does. Developing language users are, in other words, dynamic meaning makers. They can be seen to be engaging in the process of

constructing new meanings in, and through, language use as well as constructing new, alternative ways of communicating existing meanings in the course of language development (see also Lindfors' (2008) discussion of the notions of the emergent reader, writer and conversationalist in her work on young children's oral and written language development, and how such recent re-conceptualizations mark an important shift in thinking about children's developing abilities).

In this thesis, when the term 'learner' is used in the conventional way (i.e., through a deficit lens), it is printed with quotation marks to avoid confusion of conceptualization with the more general usage of *learner* mentioned above. Of course, the established area of research known as learner corpus research, with its related terms such as 'learner corpus' and 'learner corpora', retains its name as such with no alternation.

1.2.2 *Conventional language use* versus *innovative language use*

The thesis also draws a distinction between conventional and innovative language use to describe the developing language as a meaning-making resource. In the current Applied Linguistics literature, it is not difficult to find research articles reporting on work on 'learner errors' or on the socially constructed concept of 'accuracy' in 'learner language'. The notion of accuracy, together with fluency and complexity, is, in fact, a central proficiency dimension figuring as a major research variable in applied linguistic research (see, e.g., Skehan, 1998; R. Ellis, 2003, 2008; R. Ellis & Barkhuizen, 2005; Larsen-Freeman, 2006; Housen & Kuiken, 2009). Relevant to the present discussion is that what is viewed deviant or inaccurate is considered in this thesis as instances of *innovative language use*: they are natural, equally valid features of language that show divergent, nonconforming use. *Conventional language use*, on the other hand, refers to the use of linguistic commonalities as documented in descriptive grammars and dictionaries or linguistic properties conforming to such use.

There are two important points to note here. First, although this distinction is of my own making, I recognize that it is not wholly consistent with my overall position. The division between conventional and innovative language use is an artificial one, for all instances of conventional and innovative language clearly add up to what we call, holistically, the developing language. Underlying the false dichotomy between conventional language with its 'conforming' linguistic features and innovative language with its 'nonconforming' linguistic features is the notion of 'norm', which is based on an external, idealized version of

competence as a point of reference or based on an external set of monolingual 'native-speaker' data, all of which can be called into question on scientific grounds. Essentially all linguistic features observed in the developing language are inherently instances of natural human language use, although the distinction made between linguistic conventions, on the one hand, and innovations, on the other, might be a good reminder of how aspects of language use have over the years been, for various social, cultural and historical reasons, associated with certain kinds of ideological views that actually have no scientific basis in their linguistic properties.

However, the distinction, though artificial, serves two purposes in this thesis. The first is that the notion of innovative language use provides a convenient means, in the absence of, for example, an existing grammatical terminology, of describing an equally valid but nonconforming language feature in the developing language. The second is that the distinction between conventional and innovative language use helps to sharpen the analysis by making possible the observation of whether and to what extent the developing language is approaching more, or less, conventional use over time, and what this convergence, or divergence, of changing language use might mean in the study of human language development.

Second, I hope the discussion thus far has made clear that aspects of conventional and innovative language use with associated meanings are not fixed and static entities; they interact and change in the course of language development. As will be apparent in the findings of the series of studies to be reported in this thesis, some aspects of the developing language are becoming conventionalized over time while others are observed to be moving away from the convention, signifying a living, dynamic language at work.

To sum up, a number of terms are distinguished in this thesis. The term *learner* is reserved in the discussion of an individual or individuals as agents of learning in general, such as 'a life-long learner' or 'we are all learners in some areas of our language'. The term *developing language users*, on the other hand, is used here to refer to more specific, meaning-making individuals or, as noted above, the meaning makers, with continual interpretation and reinterpretation of experiences *in* and *through* language use by ways of formal study of a second language or informal engagement in the process of language learning; arguably, we are all language users in development, with diverse, growing experience in and through meaningful language use.

This second term signifies and acknowledges a more empowering view of individuals than the more familiar but potentially misleading conception of 'learner' (note the quotation marks): it emphasizes individual agency and creativity. Developing language users, such as the secondary school students or ELUs considered in this thesis, are seen as individuals actively engaging in the deployment and reconstruction of language resources at hand in their own meaning-making practices and in language development, rather than as 'learners' having to rectify 'errors' and imitate 'native speakers'. The construct of a developing language is, accordingly, the preferred term over 'learner language'.

A distinction has also been made between conventional and innovative language use. *Conventional language use* refers to the use of linguistic commonalities attested in descriptive grammars and dictionaries or linguistic properties conforming to such use, while *innovative language use* refers to the use of nonconforming but equally valid features of natural language. Both are two sides of the same coin, together making up the whole developing language. Language development over time is interpreted here on the basis of whether and to what extent specific language features, patterns and associated meanings in the developing language conform to conventional language use as time goes on, or are moving away from it towards innovative language use.

1.3 Background of the thesis

The study of human language and human language development has been a fascinating and rich area of scientific enquiry for centuries. While the focus of work by early biblical scholars in the thirteenth century may not really be considered 'the study of human language', an emphasis on the analysis and use of real language data has been central to research by lexicographers such as Samuel Johnson in 1755 and grammarians such as Otto Jespersen in 1909–1949 (see reviews by, e.g., Francis, 1992; Leech, 1992; Kennedy, 1998, Chapter 2; Stubbs, 2004; McCarthy & O'Keefe, 2010; Hunston, 2012). This empirical approach to language study, perhaps by necessity, underlines much of the research on language development. One of the earliest records of research of this kind is by Charles Darwin. Darwin (1877) kept a diary with notes based on close observation of how one of his infants developed his means of communication, among other things, over time. There continue to be a good number of diary studies observed over the years (see Behrens, 2008, for a review), including a similar longitudinal study conducted a century later by M. A. K. Halliday (1975),

who studied his son's language development by noting with pencil and notebook meaningful expressions the latter produced.

This tradition of empirical investigation, to which the research to be reported on in this thesis belongs, is often contrasted with the tradition of Chomskyan linguistics. While the first half of the last century saw important work on collocations and vocabulary control produced by pioneers in English language teaching (ELT) such as Palmer (1933), Thorndike and Lorge (1944) and West (1953) as well as on grammar by Fries (1952), all of which based on the analysis of collections of naturally occurring texts, this emphasis on performance data shifted to a focus on introspective data in the second half of the century for many years. This is generally known as the Chomskyan revolution (see, e.g., Davies & Elder, 2004). For Chomsky (e.g., 1957, 1965), competence, in the form of mental knowledge possessed by an ideal native speaker, is prioritized over performance, which is described as being characterized by a number of limitations.

The influence of Chomskyan linguistics is still strongly felt today in some areas of research, especially in the study of first and second language acquisition (e.g., Clahsen, 1996; White, 2003, 2013; see also Isac & Reiss, 2013). Relevant to the present thesis is how Second Language Acquisition (SLA) as a field approaches the study of second language development. Perhaps not surprisingly, two research traditions have been suggested to characterize SLA: one is known as the formalist approach, influenced by Chomsky's notion of an innate Universal Grammar, while the other is known as the interlanguage approach, influenced by Selinker's (1972) seminal paper bearing the same title (Ortega, 2009). The formalist approach places emphasis on language competence over language development, with methodology being focused on grammatical judgements and experimental data to explore the underlying linguistic competence of the developing language user. The interlanguage approach, on the other hand, places emphasis on development over competence, with the methodology relying on experimental and free production data (see Ortega, 2009, Chapter 6, for more relevant discussion). While, as will be suggested in Chapter 2, the conception of interlanguage as initially introduced in 1972 differs to some degree from my view of the developing language, the focus of the interlanguage approach on performance data is consistent with an empirical approach to language study discussed above, which is the methodology employed for the series of studies to be reported in this thesis.

Specifically, the studies to be presented in the next few chapters draw upon an inductive methodology closely associated with a school of corpus linguistics led by John Sinclair. This is a lexically oriented methodology, also known as a corpus-driven approach (Tognini-Bonelli, 2001) or a word-based approach (Hunston, 2002). It focuses on the study of the behaviour of individual words, supported by computer analysis of their frequencies in a corpus of texts and the immediate co-text of these words. It may be important to point out that together with a body of research of sociolinguistic and functional orientations (e.g., Hymes, 1972; Labov, 1972, Labov & Harris, 1994; Halliday, 1978), the work of Sinclair (e.g., 1991, 2004) and his associates (e.g., Francis, 1993; Carter & McCarthy, 1999; Hunston & Francis, 1999; Hunston, 2003, 2006; Groom, 2010) directly challenges Chomsky's idea of a reliance on the linguistic intuition of the native speaker and prioritizes performance data as empirical evidence of language use over Chomsky's preferred competence. The oft-quoted 'one does not study all of botany by making artificial flowers' from his (1991, p. 6) text underlies Sinclair's empirical view of language study, and his dictum to 'trust the text', the title of his paper first published in 1992 and later the title of his (2004) volume, exemplifies his long-term position and commitment to evidence-based language research.

It may also be necessary to note that while this thesis follows this tradition of corpus approach (counting words, looking at comparative frequency, examining the immediate co-text of specific words), the research reported herein differs from past learner corpus research, with its historical roots in corpus linguistics, in two significant respects. First, the research is concerned with the development of a second language *over time*, with a series of studies to be conducted based on data by the same writers at different points in time. It should be pointed out that there have been important studies of second language development using longitudinal data. For instance, Larsen-Freeman (2006) examined the written production of five students of English who performed the same task at four different points in time over a six-month period; and Klein and Perdue (e.g., 1992, 1997) studied 40 adult immigrants based on their oral production data through different sources, from free conversations to more controlled discourse elicitation tasks, with each task repeated at least three times over a period of about 30 months. These studies are, however, generally not considered as research conducted within the corpus linguistic tradition: the data in Larsen-Freeman (2006), for example, were approached from a complex, dynamic system perspective while in Klein and Perdue (1992, 1997), a functionalist approach.

The longitudinal corpus to be used in this thesis, while in no way comparable to the ambitious and impressive project associated with Klein and Perdue, is unique on its own in that it may be the largest of its kind in the field on the ground of homogeneity of data with the number of developing language users involved. It consists of four sub-corpora comprising 496 narrative texts in total, written by 124 ELUs based on one single, repeated task. They were collected at four different points in time over a 24-month period, thus providing an invaluable focused and homogenous resource for the study of the developing language (see Chapter 3 for more details).

Second, unlike much previous learner corpus research, language development is studied here based on internal, rather than external, comparisons. There is no external, reference corpus used and there are no external benchmarks to measure the developing language. The internal workings of the developing language itself are the standards for comparison.

These two important points are demonstrative of my endeavour, to paraphrase Jespersen (1909-1949, p. v), to represent the developing language not as a set of stiff dogmatic precepts, according to which some things are correct and others absolutely wrong, but as something living and developing under continual fluctuations and undulations, something dynamic and complex, something that is essentially, in one word, human. They also reflect a commitment to an entirely new set of values attached to the traditionally conceptualized 'learner language' and ways of treating, approaching and studying the developing language in its own right.

There are three final notes to make before we turn to the next section on research questions to be addressed in this thesis. First, the work to be reported in this thesis consists of two kinds of research. In addition to a corpus approach, as noted above, I also take a more 'discourse' or 'textual' approach to data analysis in Part III of this thesis (i.e., in Chapter 8), considering how the structure of narratives constructed by selected individual ELUs changes over time. That is, as will be discussed further below, this thesis reports the findings of a total of five empirical studies. There are three corpus studies to be considered in Part II and two studies of individual narrative texts in Part III. Part III begins with a quantitative study of changes in text length over time for individual texts, followed by a narrative-discourse study examining the change, if any, in the construction of the structure of the narrative by two developing language users as time goes by. Again, for this latter study, the study of change in text production practice is based on change within, rather than between, individual ELUs.

Second, recent years have witnessed an emerging interest as well as discussion calling for a consideration of longitudinal changes within groups and cross-sectional trends between groups of language users to investigate how well they coincide and complement one another in research (e.g., Jarvis & Pavlenko, 2008; Byrnes et al., 2010). To this end, insights are drawn from the frequency analysis of another corpus of a cross-sectional design, comprising texts of a comparable nature by 858 developing language users (i.e., based on the exact task and written by similar age groups of ELUs) to ascertain if similar trends in language use are observed in the two data sets (i.e., one longitudinal and the other cross-sectional) in this thesis. It is important to remember though that this thesis is primarily concerned with understanding how second language development takes place over time, with observations about internal changes in the developing language being prioritized over those based on external comparisons. The longitudinal corpus is therefore the principal resource I draw upon to make sense of the language developmental process.

Finally, a note is in order on the motivation behind the compilation of the longitudinal corpus used in this thesis, which will be identified in Chapter 3 as LoCDeLUNT (**Lo**ngitudinal **C**orpus of **De**veloping **L**anguage **U**ser **N**arrative **T**exts). I have worked with Malaysian students of English for over 10 years, mainly with students from primary, secondary and high schools, and I have always pondered upon this question: how do my students develop their language over time? This has prompted me to start collecting the student data in 2007 (to be discussed further in Chapter 3) at a school I was teaching, with the consent from both the students involved and their respective parent or guardian. In this thesis, I hope I will arrive at at least an answer to this question as well as to those listed below.

1.4 Research questions

Presented below are four questions which I attempt to seek answers to in my research. The first two questions relate to my empirical quest for understanding second language development that are based on the data available from LoCDeLUNT which I have assembled over a two-year period. The third question is both ideological and methodological, reflecting my commitment to a consideration of the developing language in its own right and assessing how far it is possible to conduct analysis of the developing language without resorting to an external norm. The fourth and final question considers what the study of a longitudinal corpus adds to the literature of learner corpus research and the broader implications of the findings of this thesis for understanding human development of an additional language.

1. How does the use of three selected closed-class words among the ELUs change over time and what does this tell us not only about those words but also about other aspects of the developing language?
2. (a) As ELUs repeat the narrative task over two years, do they write more or less?
   (b) Does the structure of the narrative remain the same or does it change? If it changes, in what way does it change?
3. Given the ideological standpoint that I approach the task of researching the developing language with respect, for it to be viewed as another form of natural, human language in its own right and not against a native-speaker norm or an external point of reference, is it possible for me, the researcher, to remain true to that standpoint while I am studying the changes in the developing language that take place over time?
4. What does the study of a longitudinal corpus add to the literature of learner corpus research and to our understanding of second language development, including second language writing development, over time?

1.5 Outline of the thesis

This thesis comprises a total of eleven chapters, which are divided into four parts. Part I consists of Introduction (Chapter 1), Literature Review (Chapter 2) and Methodology (Chapter 3). Part II consists of three corpus studies, beginning with the study of the function word *that* (Chapter 4), followed by the study of *to* (Chapter 5) and *of* (Chapter 6). Part III consists of two empirical studies, beginning with the study of text length (Chapter 7), followed by a study that examines changing narrative structures in individual texts over time by two developing language users (Chapter 8). Part IV consists of three chapters. Two of them (Chapters 9 and 10) present a discussion of the four research questions presented above while the final chapter is the Conclusion chapter of this thesis (Chapter 11). More details about each of the eleven chapters are presented below.

The first chapter provides the necessary background information of this thesis. It has described the general aim of the thesis, defined key terms and concepts, provided a brief review of the relevant literature, and presented research questions to be addressed. It also shows an outline of the thesis.

The next chapter, Chapter 2, gives an overview of the existing body of key literature on the study of 'learner language' and second language development. The chapter not only provides

background information on the research to be conducted, but also highlights the gap which the thesis aims to address. Relevant issues from other areas of Applied Linguistics are also considered.

Chapter 3 outlines the methodological considerations and challenges involved in the compilation and exploitation of the longitudinal corpus. Background information on a comparable, cross-sectional corpus is also provided. This chapter also presents a discussion of the methodological decisions in the selection of the function words for the studies to be reported in Chapters 4, 5 and 6, and in the selection of the individual narrative texts for the study to be reported in Chapter 8.

Chapters 4, 5 and 6 present the findings of a detailed, qualitative concordance analysis involving a total of 7,515 instances of use of three function words, *that, to* and *of,* with their accompanying lexical and grammatical properties observed in the data. All the findings are considered in the light of making sense of how second language development takes place over time, and they reveal a number of important, and sometimes surprising, observations about the developing language as well as the developmental process.

Chapters 7 and 8 take a different approach to the investigation. Whereas Chapter 7 reports on a study of productive language use, measured in terms of whether and to what extent the developing language users are producing longer texts over time, Chapter 8 takes the whole text as a unit of analysis, rather than the corpus. The latter chapter considers the ways in which the structure of the narrative by individual language users changes as time goes on.

Chapters 9 and 10 seek to provide answers to the research questions presented in the present chapter. They follow up the five investigations by summarizing the major themes emerging from the analysis, and by linking these themes to the central goal of this thesis: making sense of the nature and process of second language development. The chapters also consider the methods used in the research, assessing both their limitations and their underlying strengths.

The final chapter, Chapter 11, concludes the thesis by considering the theoretical, empirical, methodological and practical contributions made by this thesis to Applied Linguistics. It also suggests some productive avenues for further research.

Chapter 2

Literature Review

2.1 Introduction

This thesis, as noted in the previous chapter, is concerned with how second language development takes place over time. It falls into the general field of second language acquisition (SLA) but at the same time draws upon the methodological options from corpus linguistics. Two labels for such a research orientation are witnessed in the field. One is sometimes expressed as corpus approaches to SLA while the other, more commonly known, is learner corpus research. This thesis is, however, more comfortable with positioning itself as an applied linguistic endeavour, in seeking to understand phenomena associated with human language use and development and in contributing to the research base now widely acknowledged to be of potential in addressing language-related, real-world concerns. In doing so, it also considers the research in second language writing.

This chapter is organized into two main sections. In the first section, it reviews scholarly efforts witnessed in Applied Linguistics which involve a reconsideration of ways in approaching issues of language use and language learning. A discussion follows, in the second section, of how ways in researching the developing language differ according to 'traditions' of areas of research. Here I review relevant work spanning SLA, corpus linguistics, learner corpus research and second language writing research. Whenever relevant, I point out where my research fits in and how this thesis with the study of a longitudinal corpus might add to the literature of the respective area of research. The extent to which this might be achieved is assessed in the concluding chapter (Chapter 11) of this thesis.

2.2 The changing conceptualizations of language and 'language learner'

Over the past forty years, there have been significant changes in the theory and practice of Applied Linguistics. These changes have been driven by changes in the way we think about language use and development, and by the development of a much more empirical approach to issues in the field. Since the publication of Corder's (1967) seminal article, *The Significance of Learner's Errors*, for example, an increasing number of terms that relate to

conceptulazations of language, language use and the agent involved in the language learning task have been introduced, some overlapping in orientation with each other, while others radically new in their conceptions.

What I call the developing language in this thesis was first referred to as an 'approximative system' (Nemser, 1971) or 'transitional competence' (Corder, 1967, 1981). It is Selinker's notion of interlanguage introduced in 1972, however, that has caught the attention and gains its currency. Most introductory SLA textbooks nowadays have, for example, 'interlanguage' as an entry in the index. At the time of writing this thesis, a special symposium was organized at Teachers College, Columbia University (on 5-7 October 2012, to be precise) celebrating the 40th anniversary of the Interlanguage Hypothesis. It was an event well attended by leading scholars in the field.

As the field matures over time, new terms start to be added to the literature. While all the early terms are discussed with an orientation to treat the developing language as a linguistic system in its own right, their conceptions from an 'approximative', 'transitional' or 'inter' perspective still suggest some hybridity, or some linguistic form in between two 'real' languages or systems. In the 1990s, three important new terms are introduced. Rampton (1990, p. 98), for example, calls for a greater emphasis placed on the notion of 'expert' rather than the 'native speaker'. As he points out, expertise is 'different from identification'. Klein and Perdue (1997), on the other hand, propose the Basic Variety – a well-structured, efficient and simple form of language observed to be used among the subjects in their study. The Basic Variety is part of learner varieties, which are not

> … imperfect imitations of a 'real language' – the target language – but systems in their own right, error-free by definition, and characterized by a particular lexical repertoire and by a particular interaction of organizational principles. (Klein & Perdue, 1997, p. 308)

Another noted SLA researcher, Cook (e.g., 1992, 2008), suggests that the field as a whole look at the developing language from a multicompletence perspective. He argues that the nature of the linguistic competence between monolinguals and bilinguals is fundamentally different: multicompetence involves the whole mind of the speaker, not simply their first or their second language. This perspective assumes that someone who knows two or more languages should be considered in their own right rather than as a deficient monolingual.

For Cook, a multicompetence perspective entails setting the goal of language teaching as producing a successful language user, or what Prodromou (2003, 2008) calls a 'successful user of English' in the context of English as a Lingua Franca (ELF), rather than an imitation 'native speaker'. Such a view aligns with the ELF movement (see recent reviews in Seidlhofer, 2011, 2012; Mauranen, 2012; and Jenkins, 2013). At the same time, the fluidity of language use and language learning has led researchers such as Phipps and Gonzalez (2004) and Swain (2006) to introduce the concept of languaging: the *–ing* ending indicates the process-like nature of language and language learning, rather than as a static object to be studied (see Chapter 10 later for more discussion on this term in relation to the position of this thesis). More recently, O'Rourke and Ramallo (2013) talks about the 'new speaker' as a way to consider the role of so-called 'non-native' or 'new' speakers of minority languages and to challenge historically defined, accepted approaches to minority language maintenance, revitalization and revival.

Together with the notions of *the developing language*, *the developing language user* and *the dynamic meaning maker* introduced in this thesis, this 'pool' of terms signifies the urgency to treat the developing language and its user with equal respect, and to reject a monolingual or 'native speaker' bias.

How this might be possible when it comes to the practical task of researching the developing language remains, however, a methodological challenge. As Ortega points out, after reviewing the past four decades of SLA studies investigating the development of 'learner language':

> A great challenge will be to find ways of studying learner language in its own right, rather than as an imperfect version of the target grammar; it is yet to be seen if future innovations will help overcome teleological notions of development as convergence towards the representations assumed of an ideal monolingual native speaker. (Ortega, 2009, p. 143)

The present thesis aims to respond to this challenge and asks, in one of its four research questions, whether and to what extent this is achievable.

2.3 Paradigms and approaches to studying the developing language

This section presents a discussion of how different areas of research are for the most part associated with different traditions and approaches to exploiting language data, although I recognize that (over)generalizations are sometimes inevitably made by presenting such a review. Distinct characteristics of each tradition are highlighted, and the links between the types of the studies to be considered in this thesis and the respective tradition of research practice each is associated with are considered.

2.3.1 SLA, grammatical studies and vocabulary research

The developing language, or more widely known as interlanguage (Selinker, 1972), has been a central focus of investigation in SLA research. Studies in this field have tended to focus on the development of grammatical competence. Extensive research, for example, has been carried out on accuracy order in English grammatical morphology (e.g., Dulay & Burt, 1974; see also Bardovi-Harlig, 2000, and Goldschneider & DeKeyser's 2001 meta analysis); the developmental sequences of English verb and phrase negation (e.g., Zobl, 1980, 1982); the formation of questions and relative clauses (e.g., Schachter, 1974; Huang & Hatch, 1978; Pienemann et al., 1988; Doughty, 1991; Hu & Liu, 2007); and the developmental sequences of constituent movement and word order among learners of German (e.g., Meisel et al., 1981; R. Ellis, 1989; Pienemann, 2005) (see both Pica, 2005 and Ortega, 2009, for succinct, complementary reviews of all these areas).

At the same time there is a separate but rapidly growing strand of research on vocabulary learning (see, e.g., Schmitt & McCarthy, 1997; Carter, 1998; Read, 2000; Nation, 2001; Bogaards & Laufer, 2004; Boers & Lindstromberg, 2008; Fitzpatrick & Barfield, 2009; Milton, 2009; Schmitt, 2010). The apparent parallel developments of grammatical studies and vocabulary research are perhaps no accident: this is very much reflective of a long history of conceptualization, dating back to at least the 1500s (e.g., Musumeci, 2009; see also Howatt, 2004), which views grammar and vocabulary (or lexis) as two independent domains of language study. (Note, though, that this is of course a generalization as titles such as Schmitt & McCarthy, 1997; Carter, 1998; Nation, 2001; and Schmitt, 2010, have chapters discussing phraseological phenomena, including collocations and idioms.)

2.3.2 SLA and a deductive methodology

This has a huge impact on how a great deal of SLA (and in general, of applied linguistic) research is approached. A dominant practice has been to pre-determine an area of language feature of interest and then identify and examine the relevant instances as they occur in the developing language.

An example of this practice can be observed in much of the earlier research conducted in the 1970s, commonly known as the morpheme studies, which investigated the order of acquisition of such grammatical features as articles, *-ing*, plural *–s* and possessive *–s*. Some key studies in this area include Dulay and Burt (1973, 1974), Bailey, Madden and Krashen (1974), and Larsen-Freeman (1976). It has been proposed, for example, that the developing language user acquires progressive *–ing* first, followed by plural, copula, auxiliary, article, irregular past, regular past, 3rd person singular and finally, possessive *–s*. All this constitutes what Krashen (1977) calls the 'natural order' of acquisition regardless of age or first language background of the developing language user.

Although the morpheme studies show consistent general findings (Larsen-Freeman & Long, 1991), they have received some stringent criticisms. The main inadequacy pointed out of these studies is their failure to account for the fact that the mastery of grammatical items does not take place one at a time (Rutherford, 1988); rather the developing language user acquires numerous things imperfectly all at once (see also Nunan, 1996). Later research has focused on the acquisition of individual morphemes and syntactical structures (e.g., question formation). The body of research on syntactical structures signifies a clearer deductive approach linked to the pre-selection of an area of language feature, followed by an identification and the study of the relevant instances as they occur in the developing language. This research has shown strong evidence that the developing language user gradually and systematically passes through a series of stages in second language development. All this has been claimed as findings supporting the existence of universal patterns of development (see reviews by R. Ellis, 2008; Ortega, 2009 and Mitchell et al., 2013).

Table 2.1 Developmental stages for question formation (Spada & Lightbown, 2002)

Stage 1	**Single words, formulae or sentence fragments**	*Children?* *What's your name?* *A spot on the dog?*
Stage 2	**Declarative word order**	
	no inversion, no fronting:	*It's a monster in the right corner?* *The boys throw the shoes?*
Stage 3	**Fronting**	
	wh-fronting, no inversion:	*Where the little children are?* *What the dog are playing?*
	do-fronting:	*Do you have a shoes on your picture?* *Does in this picture there is four astronauts?*
	other-fronting	*Is the picture has two planets on top?*
Stage 4	**Inversion in *wh*- and yes/no questions**	
	copula in *wh*- questions:	*Where is the sun?*
	auxiliary other than do in yes/no questions:	*Is there a fish in the water?*
Stage 5	**Inversion in *wh*- questions**	
	inverted *wh*- questions with do:	*How do you say [proche]?*
	inverted *wh*- questions with auxiliaries other than do:	*What's the boy doing?*
Stage 6	**Complex questions**	
	question tag:	*It's better, isn't it?*
	negative question:	*Why can't you go?*
	embedded question:	*Can you tell me what the date is today?*

To gain a better appreciation of what is meant by developmental patterns based on structures, let us now consider one example of a developmental sequence for question formation. As shown in Table 2.1, the acquisition of question formation is suggested to be based on a series of stages, from single words (*Children?*), formulae (*What's your name?*) or sentence

fragments (*A spot on the dog?*) to such complex questions as question tags (*It's better*, *isn't it?*), negative questions (*Why can't you go?*) and embedded questions (*Can you tell me what the date is today?*). Spada and Lightbown (2002, p. 124) further observe the following:

> …at each stage, some of the questions learners produce may be grammatical within a particular context. Indeed, at Stage 1, chunk-learned whole questions may appear quite advanced. But this does not mean that the learner has mastered all aspects of question formation. As they progress to higher stages, they are able to manipulate more linguistic elements. Thus a Stage 3 question such as 'What the dog are playing?' may be more advanced than an apparently correct question such as 'What's your name?'.

What we have considered till this point concerns the development of language as a rule-based system. This focus on rules or grammar has indeed been observed to dominate most of the discussion and research in earlier SLA studies, with vocabulary being considered as a neglected backwater (Meara, 1980; see also Meara, 2002). Vocabulary and grammar need not be viewed as two separate entities, however; indeed over two decades of corpus research has shown how lexis and grammar regularly interact and co-occur in natural language use. Corpus research of this kind, which reveals the interdependence between lexis and grammar, is most closely associated with the work of John Sinclair and his associates (e.g., Sinclair, 1991, 2004; Francis et al., 1996, 1998; Carter & McCarthy, 1999; Hunston, 2003).

2.3.3 Corpus linguistics and an inductive methodology

While a deductive methodology, as just considered in Section 2.3.2, has been dominant in SLA research especially for such purposes as hypothesis testing, the relatively recent emergence of corpus linguistics, as represented in the work by Sinclair and his associates, has encouraged an alternative way of handling language data. This is generally known as a 'corpus-driven' approach (Tognini-Bonelli, 2001) or a word-based method (Hunston, 2002), which allows for the discovery process to be led and guided by the data. This inductive approach has proved most rewarding, or productive, particularly in uncovering how language works. It has, for example, led to a critical observation that natural language use is much more phraseological and idiomatic, or patterned, than was previously thought, which in turn has resulted in a theorization of how language works known as the idiom principle (Sinclair, 1987). Work in this area, as noted above, shows the interdependency of lexis and grammar (e.g., Francis, 1993; Francis et al., 1996, 1998; Hunston, 2003), which has challenged and

changed our traditional conceptions of a language system in which the rules are independent of specific lexical choice.

The idiom principle is part of two conceptions of language organization, with the other being the open choice principle. Each of the principles can be summarized as follows.

The idiom principle states that

> …a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments…. At its simplest, the principle of idiom can be seen in the apparently simultaneous choice of two words, for example, *of course*. This phrase operates effectively as a single word, and the word space, which is structurally bogus, may disappear in time, as we see in *maybe*, *anyway*, and *another*. (Sinclair, 1991, p. 110)

The open choice principle, on the other hand, refers to

> …a way of seeing language text as the result of a very large number of complex choices. At each point where a unit is completed (a word or a phrase or a clause), a large range of choice opens up and the only restraint is grammaticalness. (Sinclair, 1991, p. 109)

Sinclair noted that while the open choice principle is probably the most usual way of interpreting language, with virtually all grammars being constructed on this model, the idiom principle which is concerned with lexical patterning and the co-occurrence of items deserves far more attention given the presence of so many multiword sequences in the language. According to Sinclair, collocation illustrates the idiom principle, with words having their preferred phraseologies and meaning attached to the whole phrase rather than to the individual parts contained therein.

These two models of language in use are incompatible with each other, and the idiom principle has been claimed to dominate in the analysis of a text. As Sinclair (1991, p. 113) observes, 'normal text … appears to be formed by exercise of the idiom principle, with occasional switching to the open-choice principle'. These two models have further been exemplified by Hunston (2002), using the following sentence taken from a book that describes psychotherapy sessions:

> After a few moments of furious scribbling, she shifted her position, grasping the pen in her fist with the point down, much as a young child would do. (Hunston, 2002, p. 144)

Hunston (2002) studied the then 400-million-word Bank of English corpus for evidence of preferred phraseologies as employed in the sentence. Based on a statistical analysis of the corpus, she identified four multiword sequences that constitute such phraseologies which can be interpreted to be in line with the idiom principle. They are *after a few moments of*, *furious scribbling*, *shifted her position* and *much as … would*. The rest of the example sentence (e.g., *grasping the pen* and *with the point down*) seems to work according to the open choice principle as there is no evidence in the large corpus to suggest that *pen* is a collocate of *GRASP*; the word choice is arguably determined by the general grammatical rules in English given that *GRASP* is a transitive verb that takes an object, that is, *pen*. The same applies to *with the point down*.

Sinclair's work on collocation with his proposal of the idiom principle, together with his associates', can be said to popularize the study of phraseology in recent Applied Linguistics circles, although collocation as a (new) field of linguistic enquiry can be traced back to Jespersen (1924), Palmer (1925, 1933) and later Firth (1957). (See also the work by scholars of the Russian phraseological tradition, e.g., Chernuisheva, 1964, and Mel'čuk et al., 1984, 1988). Indeed we have seen an active, growing body of research devoted to the investigation and discussion of phraseology and the combinatorial potential of language.

Relevant to the present discussion is that the series of studies to be conducted and reported in Chapters 4, 5 and 6 are based on an inductive, corpus approach. These studies will explore the extent to which the second language developmental process might be captured and revealed using an inductive methodology.

A corpus approach, as it is adopted in this thesis, is essentially a quantitative approach to examining emergence of linguistic features in the developing language. It relates to the well-captured observation that 'more is different', discussed in a technical article published in the leading scientific journal, *Science* (Anderson, 1972). Fromm (2004) illustrates this point well when he discusses emergence in relation to complexity (i.e., how complexity appears and emerges in complex systems). He gave the following example:

> … when one examines a single molecule of H_2O, there is nothing that suggests liquidity:
>
> * one water molecule is not fluid,
> * one gold atom is not metallic,
> * one neuron is not conscious,
> * one amino acid is not alive,
> * one sound is not eloquent.
>
> But a collection of millions of water molecules at room temperature is clearly liquid, a collective interplay of millions of neurons produce consciousness, and a common interaction of millions of gold atoms cause metallic properties. (Fromm, 2004, p. 19)

This fully captures Anderson's (1972) 'more is different'. A similar, and perhaps also more familiar, piece of advice is noted in the study of language use: The language looks rather different when you look at a lot of it at once (Sinclair, 1991, p. 100). In essence both Anderson and Sinclair, although writing in different disciplines, highlighted the idea that quantity affects quality. For Anderson, a change of scale often leads to a qualitative change in the behaviour of a system. Likewise, according to Sinclair, examining a large body of text in a systematic way 'allows access to a quality of evidence' (1991, p. 4) that is critical for language analysis. Sinclair's quantitative analysis has, as we have seen, changed the way language is perceived: that lexis and grammar are far more interdependent on each other than were previously thought. In Sinclair's words, the 'new evidence [from large corpora] suggests that grammatical generalizations do not rest on a rigid foundation, but are the accumulation of the patterns of hundreds of individual words and phrases' (1991, p. 100).

Similarly, a quantitative approach to studying emergent properties in the developing language is an attempt to bring to the fore patterns of emergence that might have been overlooked if one had only been working with a limited set of developing language user data.

Another important feature of the research associated with this thesis, as noted in Chapter 1, is that it approaches the developing language from a longitudinal perspective. In other words, the developing language is to be examined in relation to *time*. A period of time allows for observation of how a form or structure emerges in its accompanying patterns in the developing language. Some of the best studies of second language development documented

to date are indeed those that have followed the language development of ELUs over time. As pointed out by Ortega and Iberri-Shea:

> ... it can be argued that many, if not all, fundamental problems about L2 learning that SLA researchers investigate are in part problems about "time," and that any claims about "learning" (or development, progress, improvement, change, gains, and so on) can be most meaningfully interpreted only within a full longitudinal perspective. (Ortega & Iberri-Shea, 2005, p. 26)

Using both quantitative and longitudinal approaches, the present study attempts to study how language change takes place in the developing language in relation to time. As can be seen from the discussion above, I draw upon research insights from two broad fields: SLA and corpus linguistics. They seem to be unrelated at first sight, but a recently emerging area of study is beginning to bridge the gap between these two fields. This area of study is known as learner corpus research, to which we now turn.

2.3.4 Learner corpus research

There is a growing body of research that has made use of large learner corpora to explore the developing language, thereby adding a quantitative dimension to previous, more 'traditional' SLA research. Some of the best known studies that have exploited learner corpora include Milton and Tsang (1993), Howarth (1996, 1998a, 1998b), De Cock (1998), Granger (1998), Wiktorsson (2000, 2001), Altenberg and Granger (2001), Milton (2001), Housen (2002), Cobb (2003), Tono (2004) and Nesselhauf (2005).

Perhaps the most widely known learner corpus project is the *International Corpus of Learner English* (ICLE). Coordinated and led by Sylviane Granger of the Catholic University of Louvain, this international, collaborative project consists of argumentative essays written by university students who were learners of English from a range of 16 mother tongue backgrounds (Bulgarian, Chinese, Czech, Dutch, Finnish, French, German, Italian, Japanese, Norwegian, Polish, Russian, Spanish, Swedish, Turkish and Tswana; see Granger et al., 2009). Other corpus projects include the Hong Kong University of Science and Technology learner corpus (e.g., Milton, 2001) and the Japanese EFL learner corpus (e.g., Tono, 2004). The latest addition to the list of large learner corpora is the International Corpus Network of Asian Learners of English (e.g., Ishikawa, 2011). (For a list of learner corpora available around the

world, see the website maintained by Université catholique de Louvain at http://www.uclouvain.be/en-cecl-lcworld.html.)

One of the earliest studies in this area is Granger (1998). In this comparative study of a corpus of native English writing and a corpus of writing by advanced French-speaking ELUs of English drawn from the ICLE project, Granger found, for example, that ELUs produce too few native-like prefabricated patterns and 'too many foreign-sounding ones'. Not only did the ELUs use prefabs "more as building bricks than as parts of prefabricated sections" (Granger, 1998, p. 151; see also Kjellmer, 1991, p. 124), they also employed fewer prefabs than their native-speaker counterparts and showed 'overuse' of these multiword sequences.

Similar findings were reached by De Cock et al. (1998) and Cobb (2003). In their study of a matched set of 25 transcribed French ELU and native speaker university admission interviews, De Cock et al. (1998) found that the advanced students used fewer vagueness tags (e.g., *and stuff like that*) but repeated them more frequently than native speakers. This study was replicated by Cobb (2003) who drew upon a written but comparable corpus of the developing language produced by Quebec university students. Again, it was found that the same pattern of language use was observed among advanced students, in comparison with those of their native-speaker counterparts: fewer instances and overused.

The method of data analysis associated with this line of research is known as contrastive interlanguage analysis (Granger, 1998). Studies based on this method of analysis, as can be seen, are prevalent in learner corpus research. While they are interesting and provide useful descriptions of student language, they carry with them a number of limitations. Perhaps the most serious problem with such studies lies in the well-intentioned but misguided emphasis they place on the distance between the developing language and the target, native speaker language. That is, these studies have succumbed to what Bley-Vroman (1983) first termed as the 'comparative fallacy': they failed to consider the developing language as a linguistic system in its own right (Selinker, 1972), which is claimed to be one of the few points agreed upon by all SLA researchers (Lakshmanan & Selinker, 2001).

Several studies have been conducted that can be seen to go beyond merely looking at the developing language from a comparative perspective of 'overuse' or 'underuse' involving the language of ELUs and that of native speakers. These generally relate to work by Howarth

(1996, 1998a, 1998b) and Nesselhauf (e.g., 2005). Howarth, for example, has chosen to study ELUs' multiword sequences by means of small-scale manual analysis. He looked at restricted lexical collocations (verb + noun collocations) in both native speaker academic writing and in advanced student academic writing. The student data consist of ten essays (totalling about 25,000 words) written by masters students. He concludes the study by noting that "a much greater incidence of non-standard phraseology is … found in non-native writing" (Howarth, 1998b, p. 186).

Like Howarth, Nesselhauf (2005) also hand-extracted verb + noun collocations from her learner corpus of 200,000 words. This is a German subcorpus of the ICLE. Nesselhauf drew upon dictionaries, some corpus analysis and native speaker judgements to decide on different categorizations of collocations that occurred in the student data. In her detailed analysis of these collocations checked against native speaker judgements, Nesselhauf (2005) was able to classify verb + noun sequences according to both degree of restriction and degree of acceptability. Her study further confirms findings from previous research that multiword sequences remain a serious problem in advanced student language, with L1 influence being identified as an important contributing factor (see similar comments made by Granger, 1998).

These corpus studies have provided a fuller picture of characterizations of the developing language than that presented in the research tradition established using contrastive interlanguage analysis. However, these studies, like Granger (1998), de Cock et al. (1998) and Cobb (2003), also fail to fully acknowledge the significance of dynamism in language and language learning. There is also an external point of reference involved (e.g., using native-speaker judgements), without treating the developing language as a linguistic system in its own right. That is, they challenge both the view that there is no target end state to language development (Larsen-Freeman, 2005) and the Interlanguage Hypothesis (Selinker, 1972), which postulates that 'in order to characterize the language learner's linguistic competence in the L2 accurately, interlanguage must be analysed in its own terms' (Lakshmanan & Selinker, 2001, p. 395). Studies that disregard a target language bias or a native language bias, Lakshmanan and Selinker (2001) further argue, can lead to the underestimation and/or overestimation of the developing language user's linguistic competence. An alternative approach is thus warranted for studying second language development.

One important approach to capture the relevant linguistic properties underlying the developmental process, without comparing them to those of the target language, is to adopt a longitudinal design. In such a design, one looks at the ELU's performance at different points over a period of time. This methodology is able to address several concerns raised earlier. For example, a longitudinal study essentially entails tracking language change or development over time, thus allowing one to capture the dynamism of language development. Also, this approach supports and reinforces the view that there is no fixed, homogeneous target end state to language development (Larsen-Freeman, 2005). In addition, the issue of the *comparative fallacy* (Bley-Vroman, 1983) does not arise as the developing language is now analyzed in its own terms (Selinker, 1972; Lakshmanan & Selinker, 2001).

The series of corpus studies to be presented in Chapters 4, 5 and 6 attempt to make a contribution to the existing learner corpus research particularly in this respect. As noted earlier, the studies examine second language development not only quantitatively but also longitudinally. The contrary is also true: these studies address the gap in SLA research by considering longitudinal language development by a larger number of developing language users than was available in the previous case study research.

In attempting to contribute to and advance the research base in this way, I am aware that these corpus studies raise questions to be addressed. Specifically as just suggested, there are differences in the ways SLA researchers and learner corpus researchers approach the developing language. The former group of researchers typically deals with individual ELUs over time (i.e., longitudinally) while the latter deals with large numbers of ELUs (i.e., quantitatively) but at a single point in time, using a cross-sectional design. Researching the developing language that combines both dimensions of quantity and time, rather than only quantitatively or only longitudinally, requires a different approach to thinking about and studying the developing language from that of 'traditional' SLA and learner corpus approaches; in other words, approaches and expectations might need to be adjusted. A research design based on a corpus approach raises questions about the status of the study of individuals and the study of groups, and how the two relate to each other. This concern is addressed in this thesis by studying the changing language performance among the individual ELUs through a text-based approach. As will be discussed in Chapter 3, I draw upon a matrix analysis that is developed in written discourse analysis (Hoey, 2001) and apply it in the study of second language (writing) development.

2.3.5 Second language writing research

Conscious efforts to study second language writing, as pointed out in Matsuda et al. (2009), began in the mid-twentieth century when an influx of international students was witnessed in North American higher education. Two significant events have been said to solidify the status of second language writing as a field: the introduction of the term *Second Language Writing* in Kroll (1990), and the publication of the *Journal of Second Language Writing*.

Much research in the field, however, still relies on a native-centric perspective. For example, in a recent, ambitious attempt to synthesize findings of hundreds of studies on discourse or macro properties of second language writing based on past three decades of research, Hinkel (2011, pp. 527-531) presents the following which are based on the results of comparison with the discourse structuring and ideational development observed in first language writing:

Discourse Structuring and Ideational Development in L2 Writing

Compared to L1 writers of similar social and educational backgrounds, and based on research in similar genres, L2 writers

- organize and structure discourse moves differently
- utilize discourse moves and their contents differently and inconsistently, primarily due to the negative transfer of discourse structuring conventions across various cultures
- construct or place thesis statements differently, as well as omitting them altogether
- take a logically and conceptually different approach to rhetorical development, argumentation, persuasion, and exposition/narration
- often neglect to account for counterarguments and to anticipate audience reactions
- support their arguments and claims by means of statements of personal opinions and beliefs in lieu of more substantive information
- significantly more often leave their argumentation unsupported
- sequence ideas and explanatory information differently: the norms of rhetorical structuring of discourse often do not conform to those expected in comparable written genres in English
- construct less fluent and less detailed/explanatory prose
- produce shorter and less elaborated texts
- rely more on personal opinions and include less fact-based evidence in argumentation and exposition
- over- or under-estimate the amount of readers' background knowledge and the need for textual clarity, explicitness, and specificity
- differently orient the reader to the content, as well as differently introduce and develop topics;
- delay or omit thesis/main point statements, and also omit or dramatically shorten conclusions/closings (e.g. one-sentence closings, as in: *Hopefully, scientists will find a solution to this problem soon.*)
- employ different strategies for extracting/citing information from sources, as well as paraphrasing, quoting, and including source material in their writing

- develop text cohesion differently, with weak lexical/semantic ties and theme connections, and a preponderance of overt discourse-level conjunctions
- rely on different given–new (theme–rheme) idea development
- use different sequencing, parsing, ordering, and connecting paragraph divisions, e.g., in some cases, such as those found in academic essays, L2 paragraphs need to be re-organized or divided into shorter ones, or short paragraphs need to be combined into longer ones
- differently—and often inconsistently—establish text cohesion: less frequent and less dense
- usage of cohesion devices, such as lexical, discoursal, and referential cohesive ties
- rely on repetition in order to paraphrase or establish cohesion at rates twice as high as those found in L1 writing
- develop prose that is oblique (e.g., hints) and vague (e.g., questions and allusions in lieu of direct statements)
- often take moralistic and emotionally appealing approaches to argumentation and persuasion

(Hinkel, 2011, pp. 527-528)

Similarly, limited vocabulary and grammar have been pointed out to be the most frequently cited or noted properties of second language text:

Micro Features (Grammar and Vocabulary) of L2 Writing

Compared to L1 prose, L2 texts

- exhibit less lexical variety and sophistication
- contain significantly fewer idiomatic and collocational expressions
- have smaller lexical density and lexical specificity, and more frequent vocabulary misuses
- rely on shorter sentences and clauses (aka T-units) with fewer words per clause and fewer words (e.g., nouns and modifiers) per verb
- involve high rates of incomplete or inaccurate sentences (e.g., missing sentence subjects or verbs, incomplete verb phrases, sentence fragments)
- repeat content words more often (i.e., nouns, verbs, adjectives, and adverbs)
- provide twice as many simple paraphrases or avoid paraphrasing altogether with a preponderance of referential pronouns (e.g., *this*, *that*, *it*)
- use shorter words (fewer words with two or more syllables), more conversational and high frequency words (e.g., *good*, *bad*, *ask*, *talk*)
- incorporate fewer modifying and descriptive prepositional phrases, as well as a higher rate of misused prepositions
- employ less subordination and two to three times more coordination.

L2 texts also employ

- fewer passive constructions
- fewer lexical (e.g., adjectives and adverbs) and syntactic modifiers (e.g., subordinate clauses) of sentences, nouns, and verbs

- inconsistent uses of verb tenses
- more emotive and private verbs (e.g., *believe*, *feel*, *think*)
- significantly higher rates of personal pronouns (e.g., *I*, *we*, *he*) and lower rates of impersonal/referential pronouns (e.g., *it*, *this*, *one*)
- markedly fewer of abstract and interpretive nouns, and nominalizations (e.g., *rotation*, *cognition*, *analysis*)
- fewer adverbial modifiers and adverbial clauses
- fewer epistemic and possibility hedges (e.g., *apparently*, *perhaps*) and more conversational hedges (e.g., *sort of*, *in a way*)
- more conversational intensifiers, emphatics, exaggeratives, and overstatements (e.g., *totally*, *always*, *huge*, *for sure*)
- fewer downtoners (e.g., *almost*, *hardly*)
- more lexical softening devices (e.g., *maybe*).

(Hinkel, 2011, p. 529)

Hinkel further notes that based on the research reported, the following 'error types' have been found to be 'highly common and pervasive':

Examples of Frequent Error Types in L2 Writing

The following error types have been found to be frequent in L2 writing:

- Sentence divisions, fragmented and clipped sentences, and run-ons, e.g., **So, I ask. *Sometime, one can be lack.*
- Subject and verb agreement, e.g., **Teachers of math and reading is serious about teaching.*
- Verb tenses and aspects, and verb phrases, e.g., **I remember the time when I receive a phone call from my boss that they were not satisfy with the work we've done.*
- Word-level morphology (i.e., absent or incorrect affixes) and incorrect word forms, e.g., **nation pride*, **America class is more interested than in my country*.
- Incomplete or incorrect subordinate clause structure (e.g., missing subjects, verbs and clause subordinators), e.g., **when try to be success*, **although economic is not a factor.*
- Misuses (or under-uses and over-uses) of coherence and cohesion markers, such as coordinating conjunctions and demonstrative pronouns, e.g., **At last, I completely agree with this. *The next reason is not willing to try again.*
- Singular or plural nouns and pronouns. **People want to go to school, so he work very hard on his subjects*. **The elder are given many equipments to help them in the old age.*
- Incorrect or omitted prepositions, e.g., **from my opinion*, **At some time there is this young businessman who just about takes a taxi of the airport*.
- Incorrect or omitted articles, e.g., **Finally, some people can not take good exam and telling very sad. *Some students sleep in classroom, play cellphone, play game.*
- Incorrect modal verbs, e.g., **It is also important to have adults by their side whom could advice them when they may make a mistake.*
- Spelling errors. (Hinkel, 2011, pp. 530-531)

All this is, as pointed out in Chapter 1, based on a deficit or native-centric view of the developing language. As will be shown in Chapter 8 later, the study to be presented therein considers the student writing by focusing on change within, rather than between, individual student writers, reflecting a commitment to an entirely new set of values and ways of treating, approaching and studying the developing language in its own right. It is also a direct response, noted in Section 2.2, to the challenge of finding ways of researching the developing language away from viewing it as an imperfect version of the target language.

The study also aims to contribute to the second language writing research literature with a relatively small number of studies involving secondary school students as highlighted in Leki et al. (2008), and to the growing body of research (e.g., Hanaoka & Izumi, 2012; Wigglesworth & Storch, 2012; see also reviews by Bitchener, 2012; Polio, 2012; Williams, 2012) that examines the role of the study of written language in advancing our understanding of language development.

2.4 Conclusion

From the literature review presented in this chapter, it seems clear that there appears to be a movement in the field from an initial focus on the language of the developing language user (e.g., 'approximative system', 'transitional competence', 'interlanguage') to the later emphasis on the language user (e.g., 'expert', 'successful user of English', 'the new speaker'). In between, we have a reframing practice involving 'multicompetence' and 'languaging'. All this change in the (re)conceptualization of the 'object' of our study (i.e., 'language') as well as the people we are concerned with in our research and practice, it is argued, indicates the need for finding ways to address an underlying concern in our profession: a monolingual or 'native speaker' bias and the lack of a deep and equal respect long overdue for our students, who are often perceived as 'learners' and 'defective communicators'.

The series of studies to be presented in Chapters 4 – 8 later exemplify some methodological options applied in this thesis in its attempt to rise up to the challenge of approaching and studying the developing language with respect. As reviewed in Section 2.3 of this chapter, I draw upon a lexical corpus approach and a matrix analysis as part of a general inductive methodology for this purpose. Whether and to what extent this leads to a productive and rewarding enquiry will be assessed in Chapter 9. A pedagogical model is further presented in Chapter 10 which aims to revitalize how practitioners and policymakers go about thinking

and acting to promote an approach that redresses the balance between the deficit model, within which the ‘defective communicator’ is framed, and the proposed model, which seeks to value and support developing language users as they are.

Before that, we turn to Chapter 3 for a discussion of the methodological issues considered for this thesis.

Chapter 3

Methodology

3.1 Introduction

As noted in Chapter 1, this thesis aims to make sense of how second language development takes place over time by examining the developing language of secondary school students of English as an additional language. It attempts to achieve this general goal by considering the following specific research questions:

1. How does the use of three selected closed-class words among the ELUs change over time and what does this tell us not only about those words but also about other aspects of the developing language?
2. (a) As ELUs repeat the narrative task over two years, do they write more or less?
   (b) Does the structure of the narrative remain the same or does it change? If it changes, in what way does it change?
3. Given the ideological standpoint that I approach the task of researching the developing language with respect, for it to be viewed as another form of natural, human language in its own right and not against a native-speaker norm or an external point of reference, is it possible for me, the researcher, to remain true to that standpoint while I am studying the changes in the developing language that take place over time?
4. What does the study of a longitudinal corpus add to the literature of learner corpus research and to our understanding of second language development, including second language writing development, over time?

In the present chapter, the methodology used for the research to be reported is discussed. I first present a description of the participants and the corpora to be used in the study, including information on how the data were collected and processed. A discussion follows of the analytic procedures used in this study to identify the selected words and texts for relevant analysis, and to decide on when a pattern of language use is considered to have 'emerged' in the developing language as an attempt to understand how second language development takes place over time.

3.2 Participants

The participants of the research considered in this thesis can be divided into two groups.

1. The longitudinal group

A total of 124 students contributed their texts to the development of the longitudinal corpus used in this thesis. The number was initially 149, but like most other longitudinal studies, the project experiences attrition over time. They were Secondary One (13-year-old) students when the data collection for this corpus began in May 2007. The students were all studying at the same secondary school in Malaysia.

Among the 124 students, 77 are female and 47 male. They reported speaking the Malay language as their mother tongue at home and are all studying English at school. English, in Malaysia, has been given an official status as a 'second' language, and the national language is Malay.

2. The cross-sectional group

A total of 858 students contributed their texts to the development of the cross-sectional corpus used in this thesis. This corpus is part of a larger corpus project developed by Abd. Samad et al (2002). The participants came from secondary schools in three states, that is, Penang, Pahang and Melaka. These three states represent the southern-central, the northern and the eastern parts of peninsular Malaysia. Three secondary schools were selected from each of the three states and these schools provided a minimum of 30 students each. As benchmarking was an objective of the larger research project, all the participants were selected from schools that performed well in the 1999 national standardized examinations (i.e., the *Penilaian Menengah Rendah* or PMR, and the *Sijil Pelajaran Malaysia* or SPM). See Abd. Samad et al (2002) for further details of the participants. (See also Kementerian Pendidikan Malaysia, 2000, for the English language curriculum specified for secondary school students in Malaysia, including all the participants from both longitudinal and cross-sectional groups of this research.)

3.3 Corpora and corpus construction

I draw upon two corpora to study the process of second language development in this thesis. The first corpus, as I have briefly introduced in Chapter 1, is LoCDeLUNT (**Lo**ngitudinal **C**orpus of **De**veloping **L**anguage **U**ser **N**arrative **T**exts). The second corpus, on the other hand, consists of cross-sectional data.

3.3.1 LoCDeLUNT

LoCDeLUNT is a longitudinal corpus which consists of texts contributed by the 124 students mentioned above spanning a 24-month period, and is developed by myself. The data were collected at four different points in time beginning from May 2007. The specific timeline of data collection is: May 2007, November 2007, November 2008 and June 2009. The data were collected and used for research purposes with the consent of both students and their respective parent or guardian.

The corpus consists of 496 texts or 116,399 words (tokens) in total. Each text is coded according to the student who wrote the text and when. The 'coding' range is 001-149, reflecting the initial number of students (149) who participated in the research. This coding system also includes the letters 'a', 'b', 'c' and 'd' to indicate the four different points in time the text was produced by each student and collected for this research. Thus, 001a and 002a, for example, refer to texts written by two different students ('001' and '002') and produced at the same time ('a' – which is May 2007). 001a and 001d, on the other hand, refer to two texts written by the same student ('001') and produced at two different points in time, with 'a' indicating May 2007 and 'd' indicating June 2009.

The texts were all written within an hour, under examination conditions with no access to reference materials, in response to a repeated picture-narrating task. The task requires descriptions of how a female character has accidentally fallen into a river (or lake) and the process of how she is saved from drowning (see Appendix). A sample text from the corpus, by Student 001a, is shown below:

> Hamizah and Shafira are my friends who lives next to my house. They are friendly and kind-hearted girls. Last weekend, they went to the garden to take a fresh air. They also pluck the flower beside the lake. On the same time, Azman and his friends also went to the garden. They wanted to go fishing. They brang three fishing rode and a pail.
>
> Hamizah didn't notice that there are a lot of stone. Shafira also didn't notice it and she was hit the stone. It was too pained. She can't balance her body and fell into the lake. Hamizah was shocked and panic.
>
> She tried to ask for help. She wanted to safe her friend but she can't because she don't no to swim proper. Thankful, Azman and his friend hear the voice of girl asked for help many times. Azman ran quickly, as fast as he can to safe that girl in the water. He quickly jumped into the lake and safe Shafira.
>
> Finally, Shafira was safe and they thanked Azman and his friends very much. Hamizah and Shafira gave away some money to Azman and his friends for thanked to safe Shafira. After the incident, Azman and Shafira became a best friend.

A final note is in order on the use of a repeated-task design. A concern might arise as to whether this will make differences in language performance difficult to distinguish due to task repetition from those as part of language development. Larsen-Freeman (2006) has provided a response to such a concern as follows:

> … using the same task several times was one way of dealing with the fact that 'even subtle differences in a task can affect performance profoundly' (Thelen and Corbetta 2002: 61), leaving unanswered the question of whether the subject has control over the language resources or not. I wanted to be able to look at performance variability that might be an 'important harbinger of change, or indeed the manifestation of the very process of change' (Thelen and Corbetta 2002: 61), not variable performance that could be due to differences in tasks or contexts. (Larsen-Freeman, 2006, p. 595)

LoCDeLUNT then makes a highly homogeneous corpus in terms of genre (narrative) and focus of task or topic (about a near-drowning character to be saved), constituting an invaluable resource for the study of second language development over time. It complements existing learner corpora that are largely based on a cross-sectional design. It must be noted too that the 496 texts in the corpus are all 'untreated', with the conventionally labelled 'spelling errors' kept as they are. This reflects the overall orientation of this thesis to value and respect the developing language.

3.3.2 The cross-sectional corpus

The cross-sectional corpus to be used in this thesis consists of 858 texts by three age groups of students: Primary Five or 11-year-olds (n = 293), Secondary One or 13-year-olds (n = 301), and Secondary Four or16-year-olds (n = 264). All these texts were written in response to the same narrative task given to those students who contributed to the development of LoCDeLUNT.

This corpus, as noted above, is part of a larger corpus project developed by Abd. Samad et al. (2002), known as the English of Malaysian School Students (EMAS) corpus. This EMAS corpus project outlined the following aims: (a) to establish baseline data of the English language proficiency of Malaysian students; (b) to establish benchmarks of student English language proficiency; and (c) to examine developmental patterns through data obtained from three age levels – Primary Five, Secondary One and Secondary Four.

The cross-sectional corpus consists of 167,869 words (tokens) in total. The coding procedure was initiated by Abd. Samad et al. (2002) at the beginning of data collection. Each essay was typed out as written and saved as a text file labelled with an identifying number to maintain participants' anonymity and confidentiality. Examples of the identifying numbers used include SMIS-P-f1-(01), SMIS-P-f1-(02) and SMIS-P-f1-(03).

While a comparison of findings from the analysis of both the longitudinal and cross-sectional corpora may provide converging evidence to capture with greater confidence patterns of second language development, it is perhaps important to emphasize again that this thesis is primarily concerned with understanding how second language development takes place over time, with observations about internal changes in the developing language being prioritized over those based on external comparisons. The longitudinal corpus is therefore the main resource I draw upon to seek to understand the language developmental process.

The details of the two corpora, with word tokens and types considered, are presented in Tables 3.1 and 3.2 below.

Table 3.1: LoCDeLUNT

Time	**No of students**	**No of texts**	**No of word tokens**	**No of word types**
1 (May 2007)	124	124	22,728	1,862
2 (Nov 2007)	124	124	28,853	2,408
3 (Nov 2008)	124	124	31,711	2,633
4 (June 2009)	124	124	33,107	2,614
Total	**(124)**	**496**	**116,399**	**9,517**

Table 3.2: The cross-sectional corpus

Sub-corpus	**No of students**	**No of texts**	**No of word tokens**	**No of types**
Primary Five	293	293	32,381	2,043
Secondary One	301	301	49,496	2,690
Secondary Four	264	264	85,992	4,908
Total	**858**	**858**	**167,869**	**9,641**

3.4 Analytic procedures

As noted in the previous chapter, the tools and methods in corpus linguistics and written discourse analysis are used in this thesis to gain insights into the language developmental process. As the details of the relevant procedures will be presented in the relevant chapters (Chapters 4-8) concerning ways of approaching a series of different studies considered in this thesis, this section will necessarily be brief.

3.4.1 Corpus studies

Chapters 4, 5 and 6 are corpus studies which take single words, rather than grammatical structures, as the starting point for studying the developing language. As will be shown, the single words are chosen on the basis of their pattern of use over time as observed in the frequency lists generated from the two corpora mentioned above (particularly from the longitudinal corpus).

For example, I was looking for a word that could demonstrate the notion of emergence (to be discussed shortly) quantitatively and longitudinally for the study to be reported in Chapter 4. I noted the word *that* from the frequency lists based on the two corpora. As will be shown in Chapter 4, the function word *that* is the 13th most frequent word in the Secondary Four sub-corpus. It does not feature in the top 20 words in either the Primary Five or Secondary One frequency list. A clearer changing pattern of use of *that* was observed in the frequency list based on the longitudinal data in which an increasing pattern of use of the word was found. I therefore decided to focus on this word as the starting point for investigation into the 'ever-changing', developing language.

There are two further points to note. First, in all the three corpus studies, WordSmith Tools 5 is used for generating the frequency lists and for studying patterns of use of the selected words based on concordance lines. Second, in analyzing the developing language, I draw upon *emergence* as a criterion to study second language development. Emergence, as defined in this thesis, refers to sustained use of features in the developing language over time. It is operationalized as the appearance or presence of at least *two instances* of a word with its accompanying phrase(s), structure(s) or pattern(s) *in two different contexts* of language use, across stages or periods of development (cf. Mackey, 1999). It is to be contrasted with a more 'traditional' sense of *acquisition*, which is concerned with *accuracy* at one single point in time. That is, the former focuses on the *process*, considering the appearance of both

conventional and innovative instances in the developing language over time while the latter is concerned with the outcome or product of 'accurate' use of language. This focus on process, rather than on product, is also a key point emphasized by SLA researchers. Gass and Selinker (2008, p 82), for example, have noted the importance of considering the process 'a learner goes through in acquiring a particular form'. Similar views are expressed by Ellis (2008) and Ortega (2009).

3.4.2 Analysis of individual texts

The changing language performance over time among the individual students in this thesis is considered in two separate studies to be reported in Chapter 7 and Chapter 8, respectively. These studies look at texts exclusively from the longitudinal corpus. Chapter 7 considers productivity in language use among the 124 participants of the longitudinal corpus project in terms of increase or decrease in language production using text length analysis. Specifically, words are counted between those texts written in May 2007 and those written in June 2009. This is a simple word count exercise but comparisons of language production over a two-year period among the students should yield some meaningful results.

Chapter 8 will consider an innovative application of a discourse analytical tool as a way to complement the corpus approach in researching language development. The chapter reports on a study that makes use of a matrix analysis (see Chapter 8 for details) to investigate whether and to what extent the structure of the narrative in individual texts changes over time. The matrix analysis is a somewhat complex analysis, but as will be shown in Chapter 8 later, it is a useful method for the study of narrative texts. A step-by-step explanation of how it is applied in analyzing the student text is provided in the chapter.

3.5 Conclusion

The series of studies to be presented in the next five chapters draw upon the texts written by 124 students participating in a longitudinal project and by 858 students participating in a cross-sectional project. Together they make up the two corpora to be considered for analysis. Further, two main approaches to data analysis will be exploited: one a lexis-oriented corpus approach and the other a matrix analysis. The corpus approach is used to study linguistic patterning in the developing language while the matrix analysis is used to study individual texts for signs of changing discourse structure in the narrative texts written by the students.

Chapter 4

Corpus Study 1: *that*

4.1 Introduction

This chapter reports on my first attempt to make sense of how second language development takes place over time. It presents findings from a study based on the word *that* and its patterns of use in the ELU language. As noted in the previous chapter, this is the first of a series of three studies that are motivated by observations emerging from computer-generated frequency lists based on the ELU data. This chapter looks at *that*, a word which, as will be discussed, shows an increasing pattern of use over time in both cross-sectional and longitudinal data employed for this study. The second study, whose findings are to be reported in Chapter 5, looks at *to*, a word which remains constant in terms of its relative frequency across different points of development in the longitudinal data. The third study, to be presented in Chapter 6, examines the word *of*, which shows no obvious pattern of frequency change over time in the data. Together they represent attempts at uncovering patterns of second language development from a bottom-up, data-driven approach.

This chapter begins by briefly reviewing research relevant to the study of *that*. It then presents an analysis of a total of 1,642 concordance lines containing all the instances of use of the word *that* in the ELU data. This includes a description of all the patterns of use of *that* with their functions identified in the ELU data, and of how changes in proportion of use of these patterns take place over time. The chapter concludes with a consideration of the findings for insights into the second language developmental process.

4.2 Background

As discussed in Chapter 2, corpus research has shown that language is much more patterned than previously thought. Recurrent word sequences differentially realized as chains (e.g., Stubbs & Barth 2003), lexical bundles (e.g., Biber et al., 1999), multi-word clusters (McCarthy & Carter, 2002), n-grams (e.g., Cheng et al., 2006), and semantic sequences (e.g., Hunston, 2006/2008) have been found to be prevalent in natural language use. A relevant and important finding on patterning of language use has been about the interdependency of lexis and grammar. Rather than operate independently, it has been shown that specific lexical items

and grammatical structures regularly co-occur in natural discourse (e.g., Sinclair, 1991, 2004; Francis et al., 1996, 1998; Carter & McCarthy, 1999; Hunston, 2003; Stefanowitsch & Gries, 2003; Gries & Stefanowitsch, 2004; see also work on construction from a usage-based perspective, e.g., Tomasello, 2003; Bybee, 2008; Goldberg & Casenhiser, 2008; N. Ellis & Cadierno, 2009; N. Ellis & Larsen-Freeman, 2009).

One area of language that has attracted much attention in this strand of research is *that*-clauses with their co-occurring lexical items. In two volumes of comprehensive treatments of grammar patterns, Francis et al. (1996, 1998), for example, have studied, among other patterns, the use of verbs, nouns, and adjectives with *that*-clauses. A focus on adjectives with *that*-clauses has also been the subject of study by Mindt (2011). This complementation pattern with associated lexis has also been considered from a stance or evaluation point of view in academic disciplines. Hyland and Tse (2005a), for instance, studied the use of what they call evaluative *that* constructions in research article abstracts and found that these constructions constitute an important means through which evaluative and affective meanings are realized in academic writing. In another paper, Hyland and Tse (2005b) show the extent to which postgraduate students and published writers differ in employing this structure in abstracts from postgraduate dissertations and research articles. Charles (2006), on the other hand, studies finite reporting clauses with *that*-clause complementation in theses from two contrasting disciplines, politics and materials science, and observes how academic writers differentially construct stance according to the epistemology and ideology of their own discipline.

To date, no research, to my knowledge, has considered tracking changes that take place in the process of second language development through the study of *that* and its patterns of use. That is the purpose of the present study: I set out to explore the merit of focusing on this function word in researching language development over time. I show how the distribution and patterning of this word proportionally change across different time points in the development of a second language and why research efforts may pay off to start with a lexis-oriented approach.

4.2.1 Corpora

This study makes use of two corpora of 1,354 narrative texts in total, as described in the previous chapter, to identify and investigate all the instances of use of *that*, including *that*-clauses, by 982 school students studying English as an additional language. A short description of the two corpora is as follows (see Chapter 3 for more details):

(a) A sub-corpus of The English of Malaysian School Students (EMAS) corpus

The first corpus to be used in this study is a sub-corpus of the EMAS corpus, developed by Abd. Samad et al (2002). It is cross-sectional by design, consisting of 858 texts by three age groups of ELUs: Primary Five or 11-year-olds (n = 293), Secondary One or 13-year-olds (n = 301), and Secondary Four or16-year-olds (n = 264).

(b) The Longitudinal Corpus of Developing Language User Narrative Texts (LoCDeLUNT)

The second corpus, LoCDeLUNT, is longitudinal by design, with samples of language produced by 124 students collected by myself at four different points in time over a 24-month period. The data collection process of the second corpus started in May 2007 when the ELUs were in Secondary One at around the age of 13, and ended in June 2009 when they reached 15. This corpus consists of 496 texts written by the same group of students.

All the data from these two corpora are written narratives based on the same picture prompt. It was a narrative task that requires ELUs' descriptions of how a near-drowning character was saved (see Appendix). Together both collections of data provide important linguistic information about patterns of language use by the ELUs in this research through their responses to the same narrative task.

There is one important point to note before we turn to the next section on the analysis of the frequency of *that* in the two corpora. The overarching research goal of this thesis is to make sense of how second language development takes place *over time*. The second corpus in this study, LoCDeLUNT, which consists of longitudinal data collected at four different points in time spanning a 24-month period, is therefore the primary source of language data I work with and study for insights into the language developmental process. The cross-sectional corpus, on the other hand, is exploited mainly to find out the relative frequencies of the selected function words considered in this thesis, although it is explored relatively more extensively in the present study. As mentioned in the previous chapter, the observed

frequencies in the cross-sectional data of the selected lexis will be compared with those in the longitudinal data in three corpus studies, to be reported in this and the next two chapters. Any convergent or divergent findings, through comparisons of both sets of relative frequencies observed in the two corpora, may indicate opportunities for future research, but are not addressed in this thesis.

4.2.2 Frequency of *that* in cross-sectional and longitudinal data

We now turn to Tables 4.1 and 4.2 for frequency lists generated using WordSmith Tools 5 based on the cross-sectional and longitudinal data.

What is remarkably striking about simple frequency lists like those presented in the two tables is the amount of information they reveal about the ELUs' changing frequencies of use of individual words. As can be seen from Table 4.1, the word *that* occurs only in the 16-year-olds' frequency list: it is the 13th high frequency word observed in the language of this oldest group of ELUs. It does not feature in the top 20 words in either the 11-year-olds' or the 13-year-olds' frequency list. The fact that the function word *that* was relatively more frequently employed by the oldest group as a linguistic means to help to construct their respective narratives suggests that its use might be maturity-related: *that* may be a specific word whose use increases in line with the increase in the number of years of experience with, or exposure to, language (in this case, in instructional settings).

This observation of an increase in use of *that* at a later phase of language development holds for the longitudinal data as well, as shown in Table 4.2. This function word does not feature in the 20 high frequency word list at Time 1 (it occurs as the 35th word in the full frequency list), but emerges as the 20th high frequency word at Time 2, 15th at Time 3 and 11th at Time 4. In fact, as can be seen, there is a clearer pattern of increase in the relative frequency of use of *that* observed in the longitudinal data. (It may be important to point out that the term 'Time 1' in the research reported in this thesis is by no means the very first phase which marks the beginning of second language development; clearly it is not).

Table 4.1: Comparison of word frequencies for the top 20 words in the cross-sectional data

Rank order	Primary Five (11-year-olds)			Secondary One (13-year-olds)			Secondary Four (16- year-olds)		
1	2580	7.9497%	the	3642	7.4216%	the	5464	6.4192%	the
2	1543	4.7544%	to	2101	4.2814%	to	3320	3.9004%	to
3	1336	4.1166%	and	2047	4.1713%	and	3127	3.6737%	and
4	744	2.2925%	river	1023	2.0846%	they	1717	2.0172%	was
5	665	2.0491%	they	959	1.9542%	river	1623	1.9067%	i
6	609	1.8765%	help	873	1.7790%	was	1407	1.6530%	a
7	472	1.4544%	was	785	1.5997%	a	1263	1.4838%	they
8	470	1.4482%	her	708	1.4427%	help	1170	1.3745%	river
9	412	1.2695%	for	668	1.3612%	her	1131	1.3287%	of
10	405	1.2479%	she	641	1.3062%	for	1127	1.3240%	we
11	403	1.2418%	girl	613	1.2492%	girl	1026	1.2054%	her
12	381	1.1740%	one	586	1.1941%	she	964	1.1325%	she
13	374	1.1524%	a	525	1.0698%	of	**940**	**1.1043%**	**that**
14	322	0.9922%	into	493	1.0046%	we	899	1.0562%	for
15	282	0.8689%	fishing	487	0.9924%	were	861	1.0115%	were
16	277	0.8535%	boys	486	0.9904%	fishing	747	0.8776%	help
17	276	0.8504%	in	481	0.9802%	into	716	0.8412%	my
18	268	0.8258%	flowers	453	0.9231%	i	704	0.8271%	at
19	259	0.7981%	of	400	0.8151%	flowers	698	0.8200%	in
20	241	0.7426%	i	390	0.7947%	at	612	0.7190%	girl

Table 4.2: Comparison of word frequencies for the 20 most frequent words across the longitudinal data

Rank order	Time 1 May 2007 (13-year-olds)		Time 2 Nov 2007 (13-year-olds)		Time 3 Nov 2008 (14-year-olds)		Time 4 June 2009 (15-year-olds)	
1	1741	the	2005	the	1991	the	2078	the
2	962	to	1186	to	1289	to	1399	to
3	902	and	1168	and	1161	and	1258	and
4	608	they	806	they	671	they	824	they
5	434	river	515	river	560	river	600	was
6	400	was	487	a	524	a	598	river
7	317	a	386	was	516	was	554	a
8	287	she	312	at	367	her	462	she
9	286	help	301	help	347	she	364	her
10	263	her	295	she	335	at	354	were
11	243	for	286	were	325	i	**349**	**that**
12	229	at	286	for	316	were	338	help
13	228	of	277	their	304	of	332	at
14	222	girl	276	her	299	for	308	their
15	218	were	264	girl	**288**	**that**	302	for
16	198	into	248	fishing	288	we	258	of
17	190	because	244	of	282	help	245	fishing
18	182	flowers	221	into	260	very	240	in
19	178	very	216	very	243	fishing	239	very
20	174	are	**211**	**that**	234	because	229	into

4.3 Analysis of patterns of use of *that*

This section discusses the study of *that* and its patterns of use over time. The choice of a focus on *that*, as should be apparent by now, is primarily motivated by the observation that the word occurs increasingly frequently over time in both the cross-sectional and longitudinal data. In other words, it seems to be the case that as ELUs become more mature users of the language (i.e., 'mature' in terms of the amount of time spent or experience with the language), the word *that* is used relatively more frequently among them. However, information of this kind, useful though it is, is insufficiently indicative of how the word is differentially used by ELUs at different points in time in the process of developing a second language.

I was particularly interested in exploring the various functions of *that* as it was used by the ELUs in this study of the lowest and highest ends of 'mature language use', defined here in terms of the amount of instructed learning experience the ELUs have with the language. This would allow for an investigation of the relative frequency of use of different functions of the word by the two groups of ELUs in this study. To achieve this, all the 1,642 concordance lines containing instances of use of *that* in both the cross-sectional and longitudinal data (i.e., by the 11-year-olds and 16-year-olds, and by the ELUs at Time 1 and Time 4, respectively) were considered. Every concordance line illustrating the pattern of use of *that* was examined and the different uses of the word as observed in the two corpora were noted. The resulting list of four categories of use of *that* identified through this process is presented as follows:

1. *that* as a determiner
   e.g., They rushed to *that* place and Ali quickly jumped into the river.

2. *that* as a demonstrative pronoun
   e.g., Ali, Abu and Chong were happy heard about *that.*

3. *that* as a relative pronoun
   e.g., When they arrived the hospital, they meet the doctor *that* helped to save Maria.

4. *that* as a conjunction
   e.g., All of them were very happy *that* Lucy didn't face any critical injury.

When *that* is used by the ELUs in this study as a determiner (e.g., They rushed to *that* place), the word can be observed to be functioning within a noun phrase ('that place') to identify or refer to a specific person or thing (in this case, 'place'). Other examples from the data include:

(1) The three boys, Ahmad, Abu and Ali hear *that sound*. (004a)
Tasha looks very shy when she meets Syafiq and his friends *that day*. (013a)
My friends, Ali and Abu jumped into the river to saved my life and *that girl*'s life. (030d)

When *that* is used by the ELUs in this study as a demonstrative pronoun (e.g., Ali, Abu and Chong were happy heard about *that*), it can be observed to be occupying the position of a noun phrase ('that') to identify or refer to a specific person or thing. Other examples from the data include:

(2) A few minute later, the sound was very clearly and all of we can hear *that*. (099a)
We accepted *that* and went home together. (101a)
After *that*, Ali, Abu, Raju and Siti brought Amira to her house. (122d)

When *that* is used by the ELUs in this study as a relative pronoun (e.g., When they arrived the hospital, they meet the doctor *that* helped to save Maria), it can be observed to be serving the role of introducing a clause (underlined), especially one essential to identification. Other examples from the data include:

(3) There were many types of flowers *that* were very beautiful at the lake. (018a)
After arrive at the river, they tried to found the place *that* be suitable for they go to fishing. (077a)
They was really shocked when they knew one of the girl *that* they saw a few minutes ago had fallen to the beautiful river. (041d)

When *that* is used by the ELUs in this study as a conjunction (e.g., All of them were very happy *that* Lucy didn't face any critical injury), it can be observed to be serving the role of introducing a clause (underlined), especially one expressing a statement or reporting what someone says, thinks or feels. Other examples from the data include:

(4) They said *that* they saved Manie because she was their friend, not because of the money. (105a)
When they reached, they were shocked *that* one of the girls was fall down into the pond when she was plucking flowers. (045d)
I went towards the river bank without knowing *that* I was standing on the rough part of the river bank. (103d)

4.3.1 Complexity in patterns of use of *that*

Overall two broader groupings of use of *that*, signifying different kinds of structural complexity, may be noted from the four categories of use identified above. The first concerns usage within phrasal units. The categories of *that* as a determiner and *that* as a demonstrative pronoun exemplify this. The former functions within a phrase (e.g., They rushed to *that* place), occurring with and combining a noun to form a noun phrase, while the latter occupies the position of a noun phrase (e.g., Ali, Abu and Chong were happy heard about *that*), forming a phrase on its own.

The second may be said to be relatively more complex in structure and involves the linking of clausal units. The categories of *that* as a relative pronoun and *that* as a conjunction exemplify this. While the former introduces a relative clause that characterizes the meaning of the noun phrase it modifies (e.g., When they arrived the hospital, they meet the doctor *that* helped to save Maria), the latter introduces a complement clause that completes the meaning of a noun phrase, a verb or as in the example that follows, an adjective that precedes it (All of them were very happy *that* Lucy didn't face any critical injury).

We might expect then that developmentally, phrasal *that* constructions are used more frequently than clausal *that* constructions at an early point of language development. At a later point in time, the latter constructions might be expected to feature more prominently in language use. The extent to which this holds true empirically is considered below.

4.3.2 Analysis of phrasal and clausal *that* constructions

This section presents the analysis of the two broader classifications of use of *that* observed in the ELU data: phrasal *that* and clausal *that* constructions. A manual analysis based on 1,642 concordance lines containing all the instances of use of *that* in the two corpora in this study yielded the following results.

For the cross-sectional data:

1. As far as the 11-year-old ELU language use is concerned, a total of 229 instances of use of *that* are observed. 126 of these instances are found to be phrasal *that* while the remaining 103 are clausal *that*; and

2. As far as the 16-year-old ELU language use is concerned, a total of 940 instances of use of *that* are observed. 398 of these instances are found to be phrasal *that* while the remaining 542 are clausal *that*.

For the longitudinal data:

1. As far as the ELU language use at Time 1 is concerned, a total of 124 instances of use of *that* are observed. 72 of these instances are found to be phrasal *that* while the remaining 52 are clausal *that*; and
2. As far as the ELU language use at Time 4 is concerned, a total of 349 instances of use of *that* are observed. 150 of these instances are found to be phrasal *that* while the remaining 199 are clausal *that*.

Let us first consider the results of the analysis of the cross-sectional data.

Table 4.3: Comparison of use of phrasal and clausal *that* constructions between 11-year-olds and 16-year-olds in the cross-sectional data (%)

	Phrasal *that*	**Clausal *that***
11-year-olds	55.02%	44.98%
16-year-olds	42.34%	57.66%

Table 4.3 presents the findings of the use of the phrasal and clausal *that* constructions by the 11-year-olds and 16-year-olds in the cross-sectional data. As can be seen, there is a proportionately greater use of the clausal *that* by the older group of ELUs (57.66%) than their younger counterparts (44.98%). While the 11-year-olds make more extensive use of phrasal *that* constructions in their narrative texts, the 16-year-olds have shown a preference for the use of clausal *that* constructions to develop and craft their narratives. It should perhaps be pointed out that phrasal *that* constructions are still important language resources to employ in the narrative discourse by the 16-year-olds, accounting for 42.34% of all the instances of use of *that* in their language, but these more mature ELUs rely to a greater extent on the use of the more complex clausal *that* constructions, which allow for expressions of details and joining of ideas essential to the development of the narrative task at hand.

Whether and to what extent these findings from the study of the cross-sectional data are generalizable to developmental trends over time in a different population of ELUs remain, as in all other research using a cross-sectional design, to be addressed. These are explored, in the present study, using the longitudinal data produced by the 124 ELUs.

Important for our purpose of seeking to understand how second language development takes place as time goes by, similar findings were reached of the use of *that* observed in the longitudinal data. As shown in Table 4.4, the ELUs initially depend more heavily on phrasal *that* constructions (58.06% at Time 1) in their writing of narrative texts. This trend shifts to a heavier use of clausal *that* constructions at Time 4, where the initial 58.06% of phrasal constructions observed decline in proportion to 42.98%, giving way to the more complex constructions (57.02%). It should be pointed out that while the size of difference of the results is rather small for both the cross-sectional and longitudinal data, the findings are noteworthy. That is, based on the converging, empirical evidence of a highly similar developmental pattern in the use of *that* constructions as revealed in both the cross-sectional and longitudinal data in this study, we may make the claim that development in English as a second language in instructional settings is likely to involve a change, or progression, from phrasal *that* constructions to clausal *that* constructions[1].

Table 4.4: Comparison of use of phrasal and clausal *that* constructions by ELUs between Time 1 and Time 4 in the longitudinal data (%)

	Phrasal *that*	**Clausal *that***
Time 1	58.06%	41.94%
Time 4	42.98%	57.02%

[1] This study has considered all instances of use of *that*, including *that*-clauses, but the analysis has excluded cases of *that* omission. Users of English are, of course, free to include (as in Example 1) or omit (as in Example 2, invented) the orthographic *that* in a complement clause without changing the semantic proposition in any way:

1. She knows *that* Susan does not know how to swim. (081d)
2. She knows Susan does not know how to swim.

Future research might address this concern and ascertain whether the omission of analysis of such cases would affect the findings in any significant ways.

4.4 Simultaneous growth of lexis with clausal *that* constructions

Closer analysis of the longitudinal data reveals something more illuminating about the second language developmental process. When there was an increase in clausal *that* constructions in the ELU language over time, there was a wider range of lexis observed in the data which appear to co-occur with the clauses. Here a lexical item is considered to have 'emerged' in the ELU data when it occurs at least two times with *that*-clauses by two or more ELUs across different points of development. We first turn to Table 4.5 for a list of lexical items (words and phrases) observed to have emerged in the ELU language at Time 1 before considering the change in quantity and quality of apparent, co-emergent lexis as time goes by.

Table 4.5: Emergent lexis with clausal *that* constructions in the ELU language at Time 1

Lexis	Structure
the flowers the place the three boys said says promised	Clausal *that* constructions

As can be seen, a total of six words and phrases (three noun phrases and three verbs) are observed in the ELU language at Time 1, which seem to occur together with clausal *that* constructions. They are *the flowers, the place, the three boys, said, says* and *promised*. Examples of their use in the longitudinal data are shown below:

(5) *The flowers* that she hold also fell into the river. (015a)
After arrive at the river, they tried to found *the place* that be suitable for they go to fishing. (077a)
Unfortunately, *the three boys* that want to fishing hear the voice for help. (096a)

(6) They *said* that they saved Manie because she was their friend, not because of the money. (105a)
The doctor *says* that Sarah must have more rest. (103a)
She *promised* that she will be careful later and Balqis thanked to her friends Diana because her helped. (076a)

The noun phrases in (5) can be observed to occur together with relative clauses with *that* while the verbs in (6) can be observed to occur together with *that*-complement clauses. The latter clauses are associated with the occurrence of a limited set of verbs, which together form

what Francis et al. (1996) term the verb + *that* pattern. That is, the three verbs observed here, s*aid, says* and *promised*, belong to this set of specific verbs.

Table 4.6: Emergent lexis with clausal *that* constructions in the ELU language at Time 4

Lexis	Structure
experience friends pond something story the boys the flower the flowers the girl the place the river the shouted the three boys them two girls voice	
found found out heard hope hoped knew knows noticed promised realize realized said she said the doctor says they saw told told her told her parents told (someone) told to to see	Clausal *that* constructions
happy shocked	

Turning now to Table 4.6 which presents the results of emergent lexis observed at Time 4, we find a surprisingly large number of lexical items emerging in the data at this later point of development. A number of observations can be made. First, there is clearly a longer list of words and phrases that have emerged in the ELU language at Time 4 than at Time 1. Second,

there is an additional, emergent category of adjectives comprising *happy* and *shocked* observed at Time 4. These two adjectives, like the verbs noted above, belong to a set of specific lexis co-occurring with *that*-complement clauses, which together form what Francis et al. (1998) term the adjective + *that* pattern. Third, in addition to single word items, more word combinations (e.g., *the river, found out, told her parents*) have emerged at Time 4 that are found to occur together with relative and complement clauses with *that*. Examples of use from the longitudinal data of the adjectives and word combinations noted are shown in (7) and (8), respectively:

(7) Samad also *happy* that he was helped that girl. (036d)
When they reached, they were *shocked* that one of the girls was fall down into the pond when she was plucking flowers. (045d)

(8) After back from school, Nabil, Akmal and Khalid are decide to go fishing at *the river* that not too far from their houses. (073d)
They *found out* that the girls were a twins. (008d)
She also *told her parents* that David Villa had saved her life. (127d)

Also worth noting are lexical items observed at Time 4 which show greater qualitative differences in language choice and use than at Time 1. For example, as just mentioned, adjectives (e.g., *happy*) combining *that*-complement clauses to convey expressions of feelings emerge only at Time 4. Also, while only concrete nouns (i.e., *the flowers, the place, the three boys*) emerge in the ELU language at Time 1, an abstract noun (i.e., *experience*) has been observed to make an appearance in this context at Time 4.

Most obvious of all, however, is the emergence of a greater range of verb types in the ELU language. At Time 1, for instance, only the use of verbs that express forms of communication (i.e., *said, says, promised*) was observed. At Time 4, however, in addition to those verbs of communication identified at Time 1, verbs that relate to the mental process (e.g., *hope, hoped, knew, knows*) and to the discovering process (e.g., *heard, found, found out, noticed, realize, realized, they saw, to see*) were observed (see Francis et al., 1996, for more details on the verb + *that* pattern). Examples of use from the longitudinal data of verbs relating to the mental process and verbs relating to the discovering process are shown in (9) and (10), respectively:

(9) She really *hope* that someone could heard her shouted. (088d)
Some of them brought some souveniers and *hoped* that the both of them were fully healed. (055d)
She *knows* that Susan does not know how to swim. (081d)

(10) Ali, Abu and Chong *heard* that someone was shouted for a help. (098d)
They so suprised as they *found* that one of the girls was drowning. (008d)
After she *realize* that Siti is drowning and cannot swim to save herself, she shout to get some help. (110d)

The results in Tables 4.5 and 4.6 are reproduced below to present a clearer comparison of the changes in emergent lexical use in the ELU language from Time 1 to Time 4.

Table 4.7: Emergent lexis with clausal *that* constructions in the ELU language at Time 1 and Time 4

	Time 1	**Time 4**
Noun	*Concrete* the flowers, the place, the three boys	*Concrete* friends, pond, something, story, the boys, the flower, the flowers, the girl, the place, the river, the shouted, the three boys, them, two girls, voice *Abstract* experience
Verb	*Communication* said, says, promised	*Communication* promised, said, she said, the doctor says, told, told her, told her parents, told (someone), told to *Mental* hope, hoped, knew, knows *Discovery* found, found out, heard, noticed, realize, realized, they saw, to see
Adjective	-	happy, shocked

All this suggests how quantitatively and qualitatively ELUs change their language use as they grow with greater experience over time. There are attested, substantial differences in the choice and use of lexical items in the developmental linguistic repertoires, as shown in Table 4.7. It may be important to point out, however, that given the increasing use and range of lexis

in the ELU language, it is perhaps surprising to note the relatively small size of difference of the results presented earlier in Tables 4.3 and 4.4 in relation to the use of *that* with its clausal functions. That is, whereas it is interesting that there is a difference in proportion in the use of phrasal and clausal *that* constructions, it is surprising that the differences were not greater. Or perhaps such results are not surprising after all: the rate of lexical growth is probably greater than that of structural development in language learning. It is likely, too, that the opportunities for extending grammatical range are constrained by the nature of the narrative prompt.

While the discussion here seems to suggest a false distinction between lexical growth and grammatical development, the evidence presented in this study points to the important observation that lexical and grammatical properties constitute co-emergent building blocks of language development. This is particularly true in the case of such observed patterns in use as verb + *that* (e.g., *found out* that the girls were a twins) and adjective + *that* (e.g., *shocked* that one of the girls was fall down into the pond when she was plucking flowers). The finding of these two emergent patterns (verb + *that* and adjective + *that*), surprisingly corresponding to those observed in a different, larger scale of corpus analysis of grammar patterns (Francis et al., 1996, 1998) arising out of the COBUILD project (Sinclair, 1987), also suggests that there are areas in the ELU language which have been gradually conventionalized over time. More explorations of words such as *that*, which allow for the study of both structural elements (e.g., complement clauses) and the specific lexis co-occurring with such structures, deserve serious attention in research that seeks to understand and explain the process of second language development.

4.5 Understanding second language development through the study of *that*

In this study, it has been found that the changing frequency of use of *that*, together with the emergence of the larger unit of clausal constructions, corresponds to language development among different age groups of ELUs and among the population of 124 ELUs studied using the longitudinal data. Further, it has been observed that a greater range of words and phrases (i.e., nouns, verbs and adjectives) emerge and occur together with clausal *that* constructions as these constructions are found to increase in frequency over time. While the relation of emergent lexical items to relative clauses with *that* is arbitrary (e.g., *flowers* that were …), forming sequences of words that may be called lexical bundles (see, e.g., Biber et al., 1999), the finding of the co-occurrence of *that*-complement clauses and associated lexis indicates

how lexis-grammar combinations form co-emergent properties essential to the development of a second language.

All this provides a fascinating window into the process of second language development over time. Observed changes that take place in language development are not confined to grammatical development alone; as shown in this study, lexical growth occurs at the same time. An important observation here is that certain lexical and grammatical properties in the developing language grow and construct dynamic interaction with each other, resulting in regularities of association that may be viewed as units of language acquisition which, ultimately, drive language development over time.

Second language development, as this study reveals, is at least in part about expansion of linguistic repertoires. A relatively basic communicative repertoire of lexical items and structures is observed at an early point of language development, and over time an expansion of repertoire is witnessed, consisting of many more interrelated linguistic choices. The findings emerging from the present study support a theory of language on the interdependency of lexis and grammar (e.g., Sinclair, 1991, 2004), and are compatible with usage-based theories of language acquisition (e.g., Tomasello, 2003; Bybee, 2008; N. Ellis & Cadierno, 2009; N. Ellis et al., 2013), whose basic tenet concerning constructions as the fundamental units of language acquisition is as follows:

> … constructions [which may comprise concrete and particular items (as in words and idioms), more abstract classes of items (as in word classes and abstract constructions), or complex combinations of concrete and abstract pieces of language (as mixed constructions)] and the particular lexical tokens that occupy them attract each other, and grammar and lexis are inseparable. (N. Ellis, 2011, p. 656)

As pointed out in Hunston (2003), word and phrasal combinations are essential building blocks of language; they are, as I have argued, also critically dynamic building ingredients for second language development, evidenced in the findings of the present study.

4.6 Conclusion

Unlike most previous SLA research, the present study has considered the use of a function word, *that*, to explore second language development. Changes in patterns of use of this word are tracked and the development of a set of interrelated lexical and grammatical resources among the ELUs has been observed. The choice of focus is entirely driven by an observation

of changing frequency of a function word. The study is thus significant in two respects. Firstly it contributes to an emergent view of language acquisition that focuses on the process of second language development over time. Secondly it contributes to a lexical (corpus-driven) view of what has been commonly labelled as 'learner language', re-considered here as 'the developing language' (see Chapter 1), because it has used an individual word, rather than a grammatical structure, to track language development.

There has recently been a call for focusing on small words in corpus research into language use (Hunston, 2006), with function words being the key items. It seems also timely to draw the attention of SLA researchers and language teaching professionals in general to the same set of small words. The word *that* as examined in this chapter, for example, reveals how its use, in terms of proportion, function and complexity, changes as second language development takes place over time. The emergence of word and phrasal combinations observed through the study of *that* further points to the merits of focusing on words of this kind to explore the interactive and dynamic nature of linguistic properties in language development. Attempts at more effective instructional strategies might also benefit from directing the student's attention to how small words like *that* are used in relation to accompanying lexical and grammatical properties. All this, incidentally, also suggests how fruitful a research journey can be when one considers data and evidence from a perspective that priorities lexis and values a longitudinal approach to tracking changes in language development.

In the next chapter, I study another function word, *to*, which, as will be apparent, also shows both intra- and inter-clausal roles in discourse. Using a different word as a starting point of research enquiry, I seek to explore whether the process of second language development is all about expansion of linguistic repertoires and to what extent the finding that lexical and grammatical resources combine to form co-emergent properties in the developing language over time, holds true in the study of patterning of use of a different language feature.

Chapter 5

Corpus Study 2: *to*

5.1 Introduction

In Chapter 4, I argued that the study of a grammatical word, *that*, provides an important way through which the second language developmental process can be fruitfully investigated. Specifically the chapter has demonstrated how language resources that comprise lexis and clauses dynamically change, both quantitatively and qualitatively, as second language development takes place over time. In this chapter, I look at another function word, *to*, in the ELU language in an attempt to gain further insight into the process of second language development.

While the selection of *that* for close examination in the previous study is entirely driven by the observation of changing frequency of the word, the present study focuses on the word *to* for just the opposite reason: its relative frequency remains constant in the frequency lists generated from both the cross-sectional and longitudinal data (see Tables 5.1a and 5.1b below). This observed performance trend stands in contrast to the behaviour of changing frequency of *that* shown in Chapter 4.

It is also important to note that while *that* is an obvious choice of focus because its frequency information invites speculations about the ELUs' expansion of linguistic repertoires over time, the frequency information about the use of *to* by the same group of language users raises the question of whether *to* may be part of those aspects of language which have been acquired at a much earlier point of language development (i.e., prior to the start of the collection of the ELU data for this research) and thus remains unchanged over time in its relative frequency, or whether such frequency data may have masked important qualitative differences of language use over time by the ELUs. That is, the word *to* also constitutes an intriguing object of study in its own right.

This chapter first presents a brief survey of the relevant research, showing why a small word such as *to* deserves attention in the study of second language development, before discussing the results of the analysis of all the 4,839 instances of use of *to* in the longitudinal data to

investigate its changing patterns of use among the ELUs. A major observation emerging from the findings of the study is that second language development is about 'letting go' of what may be deemed as holding the developing language user back as much as about a process of acquiring and accumulating more linguistic resources.

This observation, as will be apparent, forms part of *the paradox of complexity* evident in the developing language analyzed in this study. This paradox relates to the idea that the developing language is, and can be demonstrably shown to be, simultaneously getting more complex and less complex over time. As noted in Chapter 1, I draw a distinction between 'learner language' and 'the developing language', with the former being commonly regarded negatively as a deficient version, with perceived errors of different kinds, of the 'real' native speaker language, and the latter being used in this thesis to signify another form of equally natural and normal expression or representation of human language. The resulting fundamental change in the ways we look at, and approach, this form of language through the distinction specified is, I argue, significant for a scientific study of language – of any form of natural human language. As will be shown in this study, the approach to treating the developing language in its own right helps to remove the bias effect coming from retraining the analysis to only one aspect of 'learner language' (some instances of use are 'accurate' or 'acceptable' while others are not), and the removal of this bias effect is especially critical to the study of the phenomenon of complexity in the developing language. Indeed, the paradox is observable only when an equal, respectful consideration of all instances of language use by ELUs is adopted, with the resulting different dimensions of complexity observed.

A central argument of this chapter, based on the findings of the study, is that the linguistic resources the developing language user enjoys at different points in time are inherently complex in different ways throughout the course of second language development, thus challenging the very notion of language development as a gradual process with stages of development along a progression from simple to complex language use.

5.2 Background

Little research effort has been directed specifically to investigating the use(s) of the orthographic wordform *to* in the study of language development, perhaps as a result of it being a function word as opposed to a content word. This is, however, in no way suggesting that a function word such as *to* merits little attention. The study presented in Chapter 4, for

example, has demonstrated the value in considering this group of small words. Also, as reviewed in Chapter 2, there is a small but growing body of research outside the field of SLA focusing on the use of these words. In text and corpus linguistics, for instance, it has been suggested that while content words reveal the substantive content or 'aboutness' of a text, function words indicate the characteristic grammatical 'style' of the text (Scott & Tribble, 2006; for a critique of this distinction, see Groom, 2010).

Recent research also suggests that not only are function words important to style, they are also an indicator of meaning. This is an argument underlying a phraseological view of language. Hunston (2006), for example, shows that these small words are crucial to textual meaning. Because a range of other words co-occur with function words which share meanings, the presence of function words themselves indicates the presence of meaning. As Groom (2010, p. 63) explains, these words "tell us at least as much the preferred meanings of a particular discourse community as they can tell us about the preferred stylistic features associated with this community".

An obvious advantage of focusing on function words is that they provide the researcher with an abundance of data for analysis, even when the corpus under analysis is relatively small, as is the case in the present thesis. Function words are among the top high frequency words in most frequency lists, with *to* often appearing in the top ten of such word lists. In the British National Corpus (BNC), for example, *to* ranks the sixth most frequent word in this collection of language data when it serves the function of an infinitive marker, and is the ninth most frequent word when it is noted to function as a preposition (Leech et al., 2001).

These two functions of *to*, one as an infinitive marker and another as a preposition, are also documented in major reference grammars such as *A Comprehensive Grammar of the English Language* (Quirk et al., 1985), *Longman Grammar of Spoken and Written English* (Biber et al., 1999) and *Cambridge Grammar of English* (Carter & McCarthy, 2006). In monolingual 'learner' dictionaries such as the Macmillan English Dictionary for Advanced Learners (MEDAL) (2007), the Oxford Advanced Learner's Dictionary (OALD) (2010) and the Collins COBUILD Advanced Dictionary (CCAD) (2012), an additional function of adverb is noted. Together they represent three major categories of use of *to*, illustrated below with examples taken from MEDAL, OALD and CCAD:

(1)	As an infinitive marker:	It's nice *to* see you again. (MEDAL) She was determined *to* do well. (OALD) She told ministers of her decision *to* resign. (CCAD)
(2)	As a preposition:	We drove *to* Newport. (MEDAL) I walked *to* the office. (OALD) She went *to* the window and looked out. (CCAD)
(3)	As an adverb:	I pulled the door *to*. (MEDAL) Push the door *to*. (OALD) He slipped out, pulling the door *to*. (CCAD)

The use of *to* is considered not only as an interesting single-word phenomenon, however; its phrasal role, as with many other function words, is also an object of serious investigation. For instance, sequences that contain *to* are also noted in the rank frequency list of the whole BNC mentioned earlier, such as *according to, up to* and *due to* (Leech at al., 2001). Such recent corpus-based observations build on earlier work, such as Quirk et al.'s (1985) discussion of a group of 'semi-auxiliaries' which, in the authors' words:

> … consist of a set of verb idioms which express modal or aspectual meaning and which are introduced by one of the primary verbs HAVE and BE. (Quirk et al., 1985, p. 143)

The examples of semi-auxiliaries they gave include *be able to, be about to, be apt to, be bound to, be due to, be going to, be likely to, be meant to, be obliged to, be supposed to, be willing to* and *have to*. This group of sequences may be extended to include negatives such as *be unable to*.

These semi-auxiliaries have also been addressed as *phrasal modals* in Francis et al. (1996). This 1996 list includes *be able to*, *be bound to, be going to, have got to, have to, be liable to, be meant to, ought to, be supposed to, be sure to, be unable to, used to* and *would do well to*. These semi-auxiliaries are also termed *lexico-modals* in Collins (2009), which also noted the additional example of *be allowed to*. But perhaps a more relevant point relating to the use of *to* is its realization, together with a range of prepositions, as a marker of modal-like meaning (Hunston, 2011). That is, it co-occurs with other content words to form expressions that relay such ideas as obligation, necessity, possibility, desirability and volition.

This review, from a description of the roles of function words in revealing stylistic merits of text and preferred meanings of a particular discourse community, to the functions of *to*

documented in grammars and dictionaries and the observation of *to* as a marker of modal meaning, suggests that the wordform *to* not only constitutes an interesting object to study in its own right but can also act as a probe for investigating broader issues beyond the investigation of the use of *to* itself. It is surprising, therefore, that the word with its patterns of use has received little to no attention in the study of second language development.

The present study focuses specifically on *to* and sets out to address the following key issues: How does the use of *to* by the ELUs change across different points of development observed in the longitudinal data, and what does this tell us about the process of developing a second language?

5.3 An analysis of *to*: Relative frequency, functions and distribution across periods of language development

Tables 5.1a and 5.1b show the rank frequency lists of the 20 most frequent words in the longitudinal corpus and the cross-sectional corpus, respectively. As can be seen, the words *the, to, and* and *they* form the top four words at all the four points of development observed in the longitudinal data. *River, was* and *a* follow the list, occupying places 5, 6 and 7 in the rank order with a slight change of relative frequencies across the four data sets. This observation is largely reflected in the cross-sectional corpus, with the top three words being *the, to* and *and* across the three age groups of ELUs.

Table 5.1a: Comparison of word frequencies for the top 20 words in the longitudinal corpus at four points of development

Rank order	Time 1 May 2007 (13-year-olds)	Time 2 Nov 2007 (13-year-olds)	Time 3 Nov 2008 (14-year-olds)	Time 4 June 2009 (15-year-olds)
1	1741 the	2005 the	1991 the	2078 the
2	**959 to**	**1185 to**	**1289 to**	**1406 to**
3	902 and	1168 and	1161 and	1258 and
4	608 they	806 they	671 they	824 they
5	434 river	515 river	560 river	600 was
6	400 was	487 a	524 a	598 river
7	317 a	386 was	516 was	554 a
8	287 she	312 at	367 her	462 she
9	286 help	301 help	347 she	364 her
10	263 her	295 she	335 at	354 were
11	243 for	286 were	325 i	349 that
12	229 at	286 for	316 were	338 help
13	228 of	277 their	304 of	332 at
14	222 girl	276 her	299 for	308 their
15	218 were	264 girl	288 that	302 for
16	198 into	248 fishing	288 we	258 of
17	190 because	245 of	282 help	245 fishing
18	182 flowers	221 into	260 very	240 in
19	178 very	216 very	243 fishing	239 very
20	174 are	211 that	234 because	229 into

Table 5.1b: Comparison of word frequencies for the top 20 words across three data sets in the cross-sectional corpus

Rank order	Primary Five (11-year-olds)			Secondary One (13-year-olds)			Secondary Four (16- year-olds)		
1	2580	7.9497%	the	3642	7.4216%	the	5464	6.4192%	the
2	**1543**	**4.7544%**	**to**	**2101**	**4.2814%**	**to**	**3320**	**3.9004%**	**to**
3	1336	4.1166%	and	2047	4.1713%	and	3127	3.6737%	and
4	744	2.2925%	river	1023	2.0846%	they	1717	2.0172%	was
5	665	2.0491%	they	959	1.9542%	river	1623	1.9067%	i
6	609	1.8765%	help	873	1.7790%	was	1407	1.6530%	a
7	472	1.4544%	was	785	1.5997%	a	1263	1.4838%	they
8	470	1.4482%	her	708	1.4427%	help	1170	1.3745%	river
9	412	1.2695%	for	668	1.3612%	her	1131	1.3287%	of
10	405	1.2479%	she	641	1.3062%	for	1127	1.3240%	we
11	403	1.2418%	girl	613	1.2492%	girl	1026	1.2054%	her
12	381	1.1740%	one	586	1.1941%	she	964	1.1325%	she
13	374	1.1524%	a	525	1.0698%	of	940	1.1043%	that
14	322	0.9922%	into	493	1.0046%	we	899	1.0562%	for
15	282	0.8689%	fishing	487	0.9924%	were	861	1.0115%	were
16	277	0.8535%	boys	486	0.9904%	fishing	747	0.8776%	help
17	276	0.8504%	in	481	0.9802%	into	716	0.8412%	my
18	268	0.8258%	flowers	453	0.9231%	i	704	0.8271%	at
19	259	0.7981%	of	400	0.8151%	flowers	698	0.8200%	in
20	241	0.7426%	i	390	0.7947%	at	612	0.7190%	girl

As noted in the introduction, the focus of the present study is *to*. Specifically I attempt to investigate how patterns of use of *to* might change over time and what this suggests about the process of second language development. All 4839 instances of *to* in the longitudinal corpus were examined with the aim of identifying categories of use. Based on an analysis of the same number of concordance lines containing all the instances, with the results manually checked to ensure accurate classification of the relevant emergent features, a total of four categories of use of *to* were identified:

1. *to* is followed by the base form of a verb
   e.g., Cheng, Afham, Ishak, Rena and I are planning *to go* fishing at Sungai Rama-rama.
2. *to* is followed by a noun phrase
   e.g., Last month, Lina and Aina went *to a river* near their house.
3. *to* is followed by an adjective, an adverb or a determiner
   e.g., They took Siti to the hospital that not *to far*.
4. *to* is followed by
   (a) a verb ending with *-s/-ed/-en/-ing*
   e.g., He just want *to saves* Ramlah.
   or
   (b) a word which takes the form of a noun, an adjective or an adverb, to construct a clause.
   e.g., One of the boys quickly jump into the lake *to safe* Maisarah.

When *to* is used by the ELUs in this study with the base form of a verb following it, it can be observed to be serving two roles: the first introduces a clause (underlined), especially one expressing a statement or reporting what someone says, thinks or feels, as in (1), while the second introduces a clause (underlined), especially one denoting a purpose along the idea of 'in order to do something', as in (2). These two uses of *to* are collectively referred to below as an infinitive marker:

(1) Amri and Chong thought that was a good idea *to* go for fishing during school holidays. (082d)
After discussed with each other, they decided *to* return back to Alia and Aina's place. (094d)
She was very excited *to* see those flowers that were as colourful as rainbow. (103d)

(2) Ali jumped into the river *to* save the girl. (025d)
Haznik quickly get some help from the villagers *to* bring the weak Aini to the hospital.(91d)
We run as fast as we can *to* find where the sound come from. (099a)

When *to* is used by the ELUs in this study with a noun phrase following it, it can be observed to be binding the noun phrase to form a prepositional phrase, which regularly indicates direction (3) and sometimes identifies the person someone speaks to (4) or the person affected by or receiving something (5). This is referred to below as a preposition:

(3) Last Sunday, Nina and I went *to* the lake nearby our house. (101a)
We walked happily *to* the river. (053d)
They will meet everyday and go *to* school together. (090d)

(4) His brother comfort him and said *to* her that she will be alright. (010a)
While walking, Aiman talked *to* Azri and Syafiq about the girls.(048d)
They walked happily while talking *to* each other. (001d)

(5) I tell them what had happen *to* Nina while we were collecting flowers. (101a)
Zaidah tell Diana's parents what happened *to* their daughter. (051d)
My mother was very thankful *to* the boys because has save me. (085d)

When *to* is used by the ELUs in this study with an adjective, adverb or determiner following it, it can be observed to be increasing the intensity of the adjective (e.g., d*eep*), adverb (e.g., *much*) or determiner (e.g., *many*) by describing the extent to which the characteristic holds, as shown in (6). This is referred to below as an adverb:

(6) At first, Hidayah did not wanted to go there because the river was *to* deep. (007a)
After the man alright, he say he was thanked *to* much to me because safe his life. (027d)
When I has pulled her up to the shore, she was fainted because there were *to* many water in her body. (046a)

When *to* is used by the ELUs in this study with a verb ending with *–s, -ed, -en* or *-ing* following it, or with a word which takes the form of traditionally classified noun, adjective, or adverb following it, it can be observed to be serving two roles: the first introduces a clause (underlined), especially one expressing a statement or reporting what someone says, thinks or feels, as in (7), while the second introduces a clause (underlined), especially one denoting a purpose along the idea of 'in order to do something', as in (8). There is no traditional grammatical category for these uses, and they are collectively referred to below as Other clause marker:

(7) Last holiday, Ali and his friends were planning *to* went fishing at the lake near their house. (033d)
She could see Memu was barely able *to* breath. (010d)
Siti promised *to* carefully when pick something at near the river. (068a)

(8) They went there *to* plucked some flowers. (090a)
Last Sunday, Aina and Nurul went to the lake *to* plucking flowers. (113a)
After the boy, Rohan, knew that I was drowning like a fish outside water, he jumped into the river *to* safe me. (103d)

As noted in Chapter 1, a distinction is made between conventional and innovative uses of language. The categories of *to* as an infinitive marker and *to* as a preposition may be classified as part of conventional language use, while the categories of *to* as an adverb and *to* as an Other clause marker may be classified as part of innovative language use.

5.3.1 Complexity in patterns of use of *to*

Overall two broader groupings of use of *to*, signifying different kinds of structural complexity, may be noted from the four categories of use identified above. The first concerns usage within clausal units. In this study, the categories of *to* as a preposition and *to* as an adverb exemplify this. The former binds a noun phrase to form a prepositional phrase (e.g., Last month, Lina and Aina went *to a river* near their house), which is part of a larger verb + *to* + noun phrase pattern (e.g., 'went *to a river*'), while the latter occurs together with an adjective, an adverb or a determiner to form an adjective or adverb phrase (e.g., They took Siti to the hospital that not *to far*). These two categories of use will hereafter be referred to as forming intra-clausal constructions.

The second may be said to be relatively more complex in structure and involves the linking of clausal units. The categories of *to* as an infinitive marker and *to* as an Other clause marker exemplify this. As shown above, *to* is used here to introduce two types of clauses (underlined): (1) clauses of purpose and (2) clauses of complementation. A clause of purpose, as the name suggests, denotes a purpose along the idea of 'in order to do something' (e.g., Ali jumped into the river *to* save the girl) while a clause of complementation completes the meaning of a noun phrase, verb or adjective that precedes it (e.g., Cheng, Afham, Ishak, Rena and I are planning *to* go fishing at Sungai Rama-rama). These two categories of use will hereafter be referred to as forming inter-clausal constructions.

From a developmental perspective, it might be expected that intra-clausal *to* constructions, such as (9) and (10) below, are used more frequently at an early point of language development than inter-clausal *to* constructions, as in (11) and (12). At a later point in time, the latter constructions might be expected to feature more prominently in language use.

(9) Last month, Lina and Aina went *to a river* near their house. (035a)

(10) They took Siti to the hospital that not *to far*. (123b)

(11) Ali jumped into the river *to save the girl*. (025d)

(12) Cheng, Afham, Ishak, Rena and I are planning *to go fishing at Sungai Rama-rama*. (108c)

The extent to which this holds true empirically is considered below.

5.3.2 Analysis of intra-clausal and inter-clausal *to* constructions

This section presents the analysis of the two broader classifications of use of *to*, with their corresponding categories of use, observed in the ELU data: intra-clausal and inter-clausal *to* constructions. A manual analysis based on 4,839 concordance lines containing all instances of use of *to* in the longitudinal data in this study yielded the following results.

As shown in Table 5.2, there is a small increase in proportion of use of the inter-clausal *to* by the ELUs over time, from 51.3% at Time 1 to 54.5% at Time 4. This suggests that inter-clausal constructions have been a slightly more dominant area of use in the ELU language since Time 1, and this area of language appears to be increasingly used, as time goes by, to construct the narrative texts among the ELUs. It should perhaps be noted that intra-clausal constructions are still important language resources to employ in the ELUs' narrative discourse, accounting for 45.6% of all the instances of use of *to* in their language at Time 4, but the increasingly mature ELUs rely now to a greater extent on the use of the more complex inter-clausal constructions, which allow for the development, joining and crafting of ideas necessary for shaping the narrative task at hand.

Table 5.2: Distribution of categories of use of *to* in the longitudinal corpus

		Total		**Time 1**		**Time 2**		**Time 3**		**Time 4**	
		Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Intra-clausal *to*	Preposition	2278	47.1	464	48.4	583	49.2	599	46.4	632	45.0
	Adverb	19	0.4	3	0.3	3	0.2	5	0.4	8	0.6
Inter-clausal *to*	Infinitive marker	2057	42.5	360	37.5	481	40.6	564	43.8	652	46.4
	Other clause marker	485	10.0	132	13.8	118	10.0	121	9.4	114	8.1
	Totals	**4839**	**100**	**959**	**100**	**1185**	**100**	**1289**	**100**	**1406**	**100**

Another important observation from Table 5.2 concerns the use of the infinitive marker and that of the Other clause marker. As should be apparent by now, these two categories constitute interchangeable alternatives of language use in the ELU language. That is, both categories function to introduce clauses of purpose and clauses of complementation. Consider the following:

1. Clauses of purpose:
   (a) Infinitive marker
   e.g., Ali jumped into the river *to* save the girl. (025d)
   (b) Other clause marker
   e.g., After the boy, Rohan, knew that I was drowning like a fish outside water, he jumped into the river *to* safe me. (103d)

2. Clauses of complementation:
   (a) Infinitive marker
   e.g., Cheng, Afham, Ishak, Rena and I are planning *to* go fishing at Sungai Rama-rama. (108c)
   (b) Other clause marker
   e.g., Last holiday, Ali and his friends were planning *to* went fishing at the lake near their house. (033d)

The only difference between these two categories of use, as noted above, is that the infinitive marker reflects conventional language use whereas the Other clause marker indicates innovative language use. What is observed from the results of the analysis is that the use of the infinitive marker increases in proportion over time, from 37.5% at Time 1 to 46.4% at

Time 4. The Other clause marker, on the other hand, decreases from 13.8% at Time 1 to 8.1% at Time 4. Another innovative category of use, *to* as an adverb, remains less than 1.0% of all instances of use throughout the four points of development. In other words, the developing language has been observed to be getting increasingly conventionalized over time.

To summarize, two main findings are noteworthy here. First, development in English as an additional language in instructional settings appears to involve a change, or progression, from intra-clausal *to* constructions to the more complex inter-clausal *to* constructions. This is not necessarily because the developing language users are increasingly using *to* differently but because other changes are manifest through the changing use of *to* (see, e.g., the next section on the observation of an increased use of associated lexis with *to*-infinitive clauses). Second, it seems that the developing language becomes gradually conventionalized as time goes by. This is likely to be a result of the ELUs' increasing instructional experience which contributes to shaping their language development over time.

5.4 An analysis of uses of *to*: simultaneous growth in lexis and grammar captured and revealed

As noted above, changes in the developing language users' preference have been found for the uses of *to* over time, from the initial intra-clausal *to* constructions to inter-clausal *to* constructions at a later point in time. One obvious question we could ask here, especially with the findings of the previous study on *that*-clauses in mind, is what other linguistic features may also be 'selected' to co-occur with inter-clausal *to* constructions, especially with the *to*-infinitive clauses which have been shown to increase in frequency over time, and whether there are differences in the use of these co-occurring features between points of development that correspond to the observed, increased use of the clauses.

Further analysis of the concordance lines containing all the *to*-infinitive clauses in the ELU language was performed. This, as will be apparent, again reveals regularities of association between lexis and structure. Specifically, when there is an increased proportion of use of *to*-infinitive clauses found in the ELU language, a wider range of words and phrases are observed to emerge at the same time. As in the previous study of *that*, this constitutes a good case to explore the relations between lexis and grammar, and their role in second language use and development.

Tables 5.3 and 5.4 list all the words and phrases occurring together with *to*-infinitive clauses as observed at Time 1 and Time 4 of the ELU language, respectively. All these items have occurred at least twice with a *to*-infinitive clause and were used by two or more ELUs at each of the two points of development. While many words were identified with their observed occurrences next to a *to*-infinitive clause, some recurring phrases, such as *do not know how* and *advised* [somebody], were identified by examining their respective context of use shown in the concordance lines with a span of up to five words to the left of a *to*-infinitive clause.

Table 5.3: Emergent lexis with *to*-infinitive at Time 1 of ELU language

Noun groups	**Verb groups**	**Adjective and adverb groups**	**Others**
a present	asked [somebody]	good	about
boys	call [the boys]	shocked/	can
the clinic	did not know how/didn't know how	very shocked	cannot
flowers/some flowers	didn't know		to the nearest phone booth
girls	dived into the river		
her	do not know how/ don't know how		
him	does not know how/doesn't know how		
himself	go/were going		
home	help/helped [somebody]		
journey	jump		
lake	jump into the water/lake/		
people	jumped into the river/lake		
public phone	know/know what		
river	like		
them	look around		
water	planned		
way	shouted		
	swimming		
	take (flowers)		
	tried/try/trying		
	walked		
	want/do not want/wanted/wants		
	went/went there		

Immediately noticeable from Tables 5.3 and 5.4 is the difference in quantity of production of words and phrases found between Time 1 and Time 4 across the four categories of noun groups, verb groups, adjective and adverb groups, and 'others' (the latter comprises closed-class words such as *can*, as well as prepositional phrases such as *on their way*, including the fixed phrase *in order to*). For each of the categories shown, there is a longer list of items produced at Time 4 than at Time 1, with the size of proportional difference being most pronounced in the categories of verb as well as adjective and adverb groups. The developing language users can be seen to have considerably expanded their language resources over the 24- month period in this study.

Table 5.4: Emergent lexis occurring with *to*-infinitive at Time 4 of ELU language

Noun groups	Verb groups	Adjective and adverb groups	Others
a good idea	advice/advised [somebody]	able	again
ambulance	asked [somebody]	It [be] to dangerous	around
body	brought [something]	excited	as loud as she could
father	called [somebody]	very happy	in order
flower	came	It [be] hard	not
some/colourful flowers	cannot wait	ready	nothing
[somebody's] friend	decide/decided	shocked/	on their way
the girls	did not know how/didn't know how	very shocked	to the nearest public phone
help	did not know what/didn't know what	sincere	to the nearby public phone
home	do not know how/ don't know how	surprised	together
journey	does not know how		
lake	don't know what		
me	find (a good/suitable) place		
money	gave [something]		
other	go/going		
parents	had		
the pond	help/helped [somebody]		
the river	jump into the river		
success	jumped into the lake/pond/river/		
them	quickly jumped into the river		
time	knew how/knows how		
	like		
	learn how		
	love/loves		
	manage/managed		
	plan/planned/planning		
	promise/promised		
	ran/ran quickly		
	rushed		
	shouted/shouted loudly		
	started		
	suggested		
	swim/quickly swam/		
	cannot swim/could not swim/		
	swam towards [somebody]		
	tells [somebody]		
	took [quantity] minutes		
	tried/try/trying		
	walked/walking/		
	walked along the river		
	want/wanted/wants		
	went/went there/		
	went to [somebody's] house/		
	went to the hospital		

As in the previous study, the list of words and phrases above can be classified into two groups of lexical items: the first consists of items associated with clauses of complementation with *to*-infinitive. Nouns such as *idea* as in *a good idea*, verbs such as *decided*, and adjectives such as *shocked* belong to this group of specific lexis. These are co-occurring items with the complement clauses (underlined), as shown in (13). In the context of the study of second language use and development, these items signify co-selection and co-emergence of lexical and grammatical properties in the ELU language:

(13) Amri and Chong thought that was *a good idea* to go for fishing during school holidays. (082d)
They *decided* to go to the Pandan river for had some fun at there together. (089d)
Her friend, Zakiah was very *shocked* to see an incident in front of her.(092d)

The second group of lexis consists of items occurring with, and arguably emerging 'independently' of, clauses of purpose with *to*-infinitive. Nouns such as *river*, verbs such as *walked* and phrases such as *as loud as she could* belong to this group of words and phrases. They do not occur together with clauses of purpose (underlined) in the strictest sense of 'co-occurrence', as shown in (14). Their observed use, however, may signify part of a general expansion of lexical resources in the ELU language.

(14) Quickly, Iman jumped into the *river* to help the girl. (100d)
While they *walked* to go to the river, they saw Rahimah and Aminah were plucked some flowers near the river. (031d)
Ziha who does not know how to swim quickly shout *as loud as she could* to get help from other people who were there. (063d)

The use of specific lexis associated with the *to*-complement clauses in the ELU language at Time 1 and Time 4 was examined to find out the extent to which the ELUs in this study have developed co-emergent lexical and grammatical resources essential to their language development.

Table 5.5 shows the specific lexis associated with *to*-complement clauses in the ELU data identified and reproduced from Tables 5.4 and 5.5. As can be seen, the overall changes in the use of the associated lexis between Time 1 and Time 4 are considerable. They are particularly pronounced with the verb groups and the adjective and adverb groups, and least so with the 'others' category (i.e., only one phrase was produced at each point, with *about to* observed at Time 1 while *in order to* at Time 4).

Table 5.5: Lexis associated with *to*-complement clauses at Time 1 and Time 4

		Time 1	Time 4
Noun group	1.	–	The 'Thought' group a good idea
Verb groups	1.		*The 'Begin' group* started
	2.		*The 'Discover' group* learn how
	3.	*The 'Help' group* helped [somebody]	*The 'Help' group* helped [somebody]
	4.	*The 'Hope' group* do not want, want, wanted, wants	*The 'Hope' group* want, wanted, wants
	5.	*The 'Like' group* like	*The 'Like' group* like love/loves
	6.		*The 'Manage' group* manage/managed
	7.	*The 'Promise' group* planned	*The 'Promise' group* decide, decided plan, planned, planning promise, promised
	8.	*The 'Remember' group* did not know how/didn't know how do not know how/ don't know how does not know how/doesn't know how know what know didn't know	*The 'Remember' group* did not know how/didn't know how did not know what/didn't know what do not know how/ don't know how does not know how don't know what
	9.	*The 'Tell' group* asked [somebody] call [somebody]	*The 'Tell' group* advice/advised [somebody] asked [somebody] called [somebody] tells [somebody]
	10.	*The 'Try' group* tried/try/trying	*The 'Try' group* tried/try/trying
	11.		*The 'Wait' group* cannot wait
	12.		The 'Say' group suggested
	13.	The 'Other verb in phase' group help	The 'Other verb in phase' group help
	14.		Phrasal modal had to
Adjective and adverb groups	1.		*The 'Able' group* able
	2.		*The 'Adequate' group* ready

	3.	*The 'Astonished' group* shocked, very shocked	*The 'Astonished' group* shocked, very shocked surprised
	4.	*The 'Delighted' group* good	*The 'Delighted' group* excited very happy
	5.		*The 'Willing' group* sincere
	6.		*The 'Easy' and 'Difficult' group* It [be] hard
	7.		*The 'Selfish' and 'Dangerous' group* It [be] to dangerous
Others	1.	about	in order

All the lexical items above have been classified according to different meaning groups in association with their respective grammar patterns (Francis et al., 1996, 1998). It must be noted that whenever there are cases of a 'good fit' between the patterns and meaning groups observed in the developing language user data in this study and the extensive range of patterns and meaning groups observed in Francis et al. (1996, 1998) based on a different and much larger data set known as the Bank of English, that suggests conventional language use and is arguably an indication of convergence upon broadly the same human language resources.

When there are cases of a misfit, however, that suggests innovative language use and the innovative items are underlined in Table 5.5. As shown in the table, many items have been underlined, including the word sequence in the noun group category emergent at Time 4 (i.e., *a good idea*) and one meaning group in the verb group category (i.e., the 'Say' group). The introduction of the 'Say' group might need further explanation. As shown in the table, it comprises the word *suggested*, a verb which is not documented in Francis et al. (1996) under the heading of the verb + *to*-infinitive pattern. This 'Say' group is, however, observed in Francis et al. (1996) under the heading of the verb + *that* pattern, with the word *suggest* being one of the attested verbs found to be occurring with this pattern. This meaning group label is therefore borrowed and used for the observed verb + *to*-infinitive pattern co-occurring with *suggested* in the ELU data. In this context, the 'Say' group may be said to be an innovative (as opposed to 'conventional') meaning group. The same case applies to the 'Thought' group.

Returning to the analysis of the emergent lexis occurring at Time 1 and Time 4, we can make two observations important to an appreciation of how changes in language use by the ELUs take place over time. First, the number of meaning groups has increased dramatically: the verb-based meaning groups, for example, has increased to 13 at Time 4 with an additional phrasal modal (i.e., *had to*) observed, almost double the number evidenced at Time 1. Similarly, the number of adjective- and adverb-based meaning groups has notably increased from two at Time 1 to seven at Time 4. Second, a greater sophistication of language use has been observed to take place as time goes by. For example, complex patterns comprising three or more elements are evidenced only at Time 4, including the complex verb patterns exemplified in *advice/advised* + [somebody] + *to*-infinitive clause (e.g., The doctor *advised* Lucy to rest for a few days) and the adjective patterns with *it* exemplified in introductory *it* + a link verb + adjective + *to*-infinitive clause (e.g., Ali tried to escape but it was *hard* to go).

Together the use of a far more extensive range of meaning groups and the development of more complex use of patterned language witnessed at Time 4 point to the remarkable expansion of repertoires among the ELUs over time. The developing users considered here can be observed to have evidently strengthened and advanced their communicative resources for meaning-making.

Overall we find changing patterns of deployment of lexical and grammatical resources among the ELUs in this study as time goes by. As has been shown, there is a proportional increase in the use of inter-clausal *to* constructions in the developing language. When *to*-infinitive clauses were studied in greater detail for accompanying, co-occurring lexical resources, a substantial increase in the use of lexical items is observed. In other words, there is an expansion of both lexical and grammatical resources, which make up a pool of emergent linguistic properties in the ELU language.

Of particular importance is the finding that co-selection of specific lexis and *to*-complement clauses has been observed and that the range of this associated lexis expands considerably over time in the developing language. The emergence of a new category consisting of a noun group (*a good idea*) and an increase both in quantity and quality of production of the categories of verb as well as the adjective and adverb groups were all observed at a later point of development. They have been found to regularly co-occur with *to*-complement clauses to form the larger combination pattern of lexis + *to*-complement clauses. While the ELUs still

used at Time 4 recurring expressions (e.g., *tried to, asked somebody to* and *shocked to*) as observed at Time 1, this range of expressions has extended and expanded to include complex patterns such as *advised somebody to* and *it* + be + *hard to*. All this richer and increasingly sophisticated language use indicates that the developing language is becoming more complex over time.

5.5 Observing development through uses of *to* over time: Less is more

The results presented thus far from the analysis of the four categories of use of *to* are rather 'neat': there are two categories of use of *to* that have been said to reflect conventional use (i.e., *to* as an infinitive marker and *to* as a preposition) while the other two innovative use (i.e., *to* as an adverb and *to* as an Other clause marker). A closer analysis of the data reveals a more complex picture of the developing language and provides more surprising results.

5.5.1 Uses of *to* as infinitive marker and as preposition

Considering first the analysis of Categories 1 (infinitive marker) and 2 (preposition), these two categories of use are well noted in reference grammars such as Quirk et al. (1985), Biber et al. (1999) and Carter and McCarthy (2006) as well as in dictionaries of English for advanced users such as MEDAL (2007), OALD (2010) and CCAD (2012). Examples of the ELUs' use of both categories have been shown in Section 5.3 of this chapter, reflecting conventional language use of an infinitive marker (i.e., to be followed by the base form of a verb) and a preposition (i.e., to be followed by a noun phrase).

What has not been discussed and is to be considered here, however, are instances of use of *to* as an infinitive marker by the ELUs which are not attested in major reference grammars or dictionaries. Specifically these are innovative instances co-occurring with modal verbs to form the pattern of modal + (negative +) *to*-infinitive (15), with *there* to form the pattern of noun + *to*-infinitive to indicate existence (16), rather than purpose or need (17) (the examples in (17) are included for comparison purposes), and with *know* to form the combination of *know* + *to*-infinitive (18):

(15) I jumped to the river for helped her because I *can to swim*. (047a)
Mary was afraided but she *did not to do* now. (121a)
The girls *could not to swim*. (077d)

(16) In that time, there were *three man across there to go fishing*. (079a)
There are *three boys at the lake to go fishing*. (122a)
There were also *three man to go fishing* at the same place. (123a)

(17) Yasmin … just crying for help, hoping that there were *people to help* them. (022b)
Amirah fell grateful because there were *someone to help* her. (090c)
… the doctor says that … there was *nothing to worry about*. (046d)

(18) … Linda don't *know to swim*. (024a)
She doesn't *know to swim*. (070a)
The water was very deep and the girl also didn't *know to swim*. (073a)

The use of these innovative, patterned sequences shown in (15), (16) and (18) above was evident at the initial point of development (Time 1) but was observed to show a general pattern of declining over time from fifteen instances of use at Time 1 to three instances of use at Time 4. Table 5.6 shows the distribution of these sequences across the four points of development examined in this study.

Table 5.6: Distribution of innovative sequences of use of *to*-infinitive across different points of development

	Total	**Time 1**	**Time 2**	**Time 3**	**Time 4**
MODAL + (neg +) *to*-infinitive	15	8	4	0	3
noun + *to*-infinitive with existential *there*	3	3	0	0	0
know + *to*-infinitive	4	4	0	0	0
Totals	**22**	**15**	**4**	**0**	**3**
% of all instances of *to*	**0.45**	**1.56**	**0.34**	**0.0**	**0.21**

While the number of occurrences of these innovative uses of *to*-infinitive is admittedly small (i.e., only 22 instances of them or 0.45% of all instances of *to* in the ELU data), their presence in the ELU language and the developmental pattern they exhibit (a declining one) are important. Often overlooked or ignored in traditional approaches to 'learner language', such instances of use, when given equal consideration as they are in the present study, reveal that

language development is not just about building an ever larger body of linguistic resources as indicated in the previous section. Their declining use suggests that there may be another fundamental process at work in second language development.

Consider both the pattern of modal + (negative +) *to*-infinitive and the pattern of noun + *to*-infinitive with existential *there*, for example. These are cases of unattested patterned language use (cf. Hunston & Francis, 1999). The 'conventional' view of such an observed declining trend of growth over time is that the 'learners' are making fewer 'errors' as time goes on. When regarded as equally valid, natural language patterns, however, the decline in their use reveals that the developing language users are progressively decreasing these patterns and that reduction is an essential mode of progression. As will be seen, a similar declining trend is observed with other instances as well as with one category of language use in the developing language. Put simply, how ELUs go about developing a second language involves as much reduction, control and abandonment of certain linguistic features and structures as accumulation and expansion of language resources.

Moving on to the use of prepositional *to*, again we find innovative instances within this category of use by the ELUs, as in (19) below:

(19) That girl's parent *thanked to Ali and Abu* because saved their son life. (030d)
Suddenly, my sister *fell down to* a river. (075d)
Ilman *told to Fatin Diana* that Afiq is a good swimmer. (097d)

Among this group of instances, the combination of THANK + *to* + noun is the most frequent in the ELU language, occurring 190 times in the longitudinal corpus (or about two occurrences per thousand words). This emergent language use suggests that a process of analogy may be at work here:

> … when a pattern is used with words with a particular meaning, speakers begin to use other words with a similar meaning with the same pattern, by a process of analogy, so that at any point in time, what words belong to a list is in a state of flux. (Hunston & Francis, 1999, p. 96)

We may compare the use of THANK + *to* + noun (or the use of the relatively less frequent combination of TELL + *to* + noun, for that matter) with such instances of use as 'say to' or 'speak to' observed in the ELU data, as shown below:

(20) Sarah *saids to* Syazi. (103a)
Kumala was very shy to *speak to* Syaz. (013b)
Siti *says to* the boys that her friend fell down into the river. (075b)

That is, it could be argued here that the developing language users drew upon a set of words ('thank', 'tell', 'speak' and 'say') with similar meanings and analogously used them with the same pattern: verb + *to* + noun.

The next step is a more detailed study of the THANK + *to* + noun combination in the ELU language. The occurrences of various forms of the verb lemma THANK (i.e., base form, *-s*, *-ed* and *–ing*) in the combination were counted. All instances of use of its counterpart, the THANK + noun combination, were also considered. Table 5.7 presents the results.

Table 5.7: Distribution of THANK + *to* + n and THANK + n across different points of development

	Total		**Time 1**		**Time 2**		**Time 3**		**Time 4**	
THANK	+ *to*+ n	+ n	+ *to*+ n	+ n	+ *to*+ n	+ n	+ *to*+ n	+ n	+ *to*+ n	+ n
base form	16	48	6	29	4	9	2	5	4	5
-s	38	11	13	2	9	4	11	3	5	2
-ed	135	112	33	28	22	26	37	28	43	30
-ing	1	0	1	0	0	0	0	0	0	0
Totals	**190 (53%)**	**171 (47%)**	**53 (47%)**	**59 (53%)**	**35 (47%)**	**39 (53%)**	**50 (58%)**	**36 (42%)**	**52 (58%)**	**37 (42%)**

As can be seen, an interesting and clearly patterned dynamic process is taking place in which THANK + noun (53%) was initially used with greater proportion than THANK + *to* + noun (47%) at Time 1 and Time 2, respectively. The latter, however, became a more frequent combination of use at later points of development. At Time 3 and Time 4, it was found that THANK + *to* + noun was used proportionally more frequently than THANK + noun (58% versus 42%, respectively, at both points of development).

That is, the former and the latter pattern can be viewed as being in competition, with the innovative use of THANK + *to* + noun found to be the preferred choice over time. The observation of the preferred THANK + *to* + noun combination at a later point of development in the ELU language signals the emergence of a seemingly lasting feature. It would be

interesting to know, although clearly beyond the scope of this study, if this combination remains in the ELU language and thus diverges from the conventional and forms a distinct feature of an emergent variety of English in the long term.

When calculated against all instances of use of *to*, however, we find a decreasing pattern of growth of the innovative use of THANK + *to* + noun over time, from 5.53% at Time 1 to 3.70% at Time 4, again suggesting a reduction process at work in second language development.

5.5.2 Uses of *to* as adverb and as Other clause marker

Turning now to Categories 3 (adverb) and 4 (Other dependent clause marker), we find equally interesting observations. The wordform *to*, as noted in Section 5.3, has been used to modify an adjective, another adverb or a determiner in the ELU data. Unlike the way a *to* adverb is illustrated in dictionaries (e.g., *I pulled the door to*, MEDAL, 2007), however, the ELUs' use of *to* suggests the exploitation of sound analogy – placing words with a similar sound in the same pattern. The conventional label for such use may be 'misspelling' (but against whose norm?). Compare the following instances of use of *to* (21) with those of *too* (22), all from the ELU data:

(21) Mah was so excited and did not realized that she was *to near* the shore. (043b)
They took Siti to the hospital that not *to far*. (123b)
While she plucked flower, she found a very nice flower but its *to close* to the lake. (083c)

(22) But you can't pick the flowers because it's *too near* to the river. (107c)
Luckily, there were three boys not *too far* from Damia and Dania. (018c)
Even her parents warned her not to get *too close* to the river, but she still wanted the flower. (007b)

All the instances of *to* classified in Category 3 (adverb) are similar instances of use exemplified in (21). These are innovative features in the developing language which may, over time, be conventionalized. As shown earlier in Table 5.2, this category of use remains less than 1.0% of all the occurrences of *to* throughout the four points of development in this study.

The fourth and final category of use of *to* is the Other clause marker category. This whole category of language use is, like the innovative use of patterns and combinations examined above, not documented in any detail in major reference grammars and dictionaries. As shown

with examples before, it comprises two main groups of features. The first is the use of *to* followed by a verb ending with *–ed, –en*, *–ing, –s* or with an irregular form. More examples from the ELU data are as follows:

(23) Shakila asked Farisha *to joined* her picked up some flower to bring home. (047c)
The girl who felt into the river was did not know how *to swimming*. (077c)
They loved *to went* there because the lake was so beautiful. (116a)

The second group consists of instances of use of *to* followed by a word that takes the traditionally perceived form of an adjective, an adverb or a noun to form a clause. More examples from the ELU data are as follows:

(24) Tina was promised *to more carefull* went at the side of river. (121a)
Finally, Fauzana was manage *to still alive*. (058d)
I told to Rahim that he have became a hero today and he just smiled while tired *to breath*. (026d)

This innovative category highlights complexity of language use evidenced even at an early point of language development. Their observed use is important, not least because their occurrences account for 13.8% of all instances of use of *to* at Time 1 in the ELU data. This category undergoes the same reduction process in second language development (the number of their occurrences was found to decrease to 8.0% at Time 4), reaffirming the idea that language development is not just about accumulating a wealth of linguistic resources.

Table 5.8 sums up the discussion of the findings on the characteristics and trends of growth of the various uses of *to* examined in this study.

Table 5.8: Uses of *to* observed in the longitudinal data

Use at Time 1	Conventional use as attested in grammar books and/or dictionaries?	Trend of growth	Use at Time 4
verb + *to*-infinitive	Yes	Increasing	Yes
MODAL + (neg +) + *to*-infinitive	No	Decreasing	Yes
noun + *to*-infinitive with existential *there*	No	Decreasing	**No**
know + *to*-infinitive	No	Decreasing	**No**
to + noun to construct a phrase	Yes	Decreasing	Yes
THANK + *to* + noun	No	Decreasing	Yes
to + adjective/adverb/ determiner to construct a phrase	No	Relatively constant	Yes
to + verb ending with *-ed, -en, -ing* or *-s* or with wordform of an adjective, an adverb or a noun to construct a clause	No	Decreasing	Yes

What is significant from the summary of the results presented earlier in Table 5.8 is that the developing language can be seen to be more complex at the beginning than it is at the end of the 24-month period of this study. Out of the eight patterns of use of *to* observed at Time 1, six are displaying a declining trend of growth and two of these (i.e., the pattern of noun + *to*-infinitive with existential *there* and the combination of *know* + *to*-infinitive) completely 'die out' by Time 4. That is, the ELUs started off with a greater number of different patterns of use of *to* at Time 1 than at Time 4, reducing the use of some and doing away with others along the way. There is a movement then from more complex to less complex language use over time.

5.6 Understanding second language development: The paradox of complexity

Taking into account now all the findings of the present study, we observe an intriguing paradox. From one perspective, the developing language is becoming more complex over time, with richer and more sophisticated language use observed involving the interplay between lexis and grammar (see Section 5.4). From another perspective, the developing language is becoming less complex over time with a declining number of different patterns of

language use (see Section 5.5). What is constructed in the developing language, then, is complexity and what is reduced is also complexity.

This paradox of complexity is observable only when all instances of language use by ELUs are considered, including those so-called deviant forms. The usual or dominant way of looking at similar findings would be to focus only on the increasing complexity, and so to see 'learner language' as becoming almost always more complex. This is because the complexity introduced by deviant forms is dismissed as simply errors, which may be seen to be decreasing. When errors are seen as decreasing, however, they are not usually described as decreasing complexity. That is, the dominant paradigm of 'learner language' analysis would be to interpret the same findings as reduction in errors and increase in complexity.

When recurring instances of innovative language use are considered as valid instances of patterns rather than just being labelled as errors and ignored in the discussion of complexity, however, we start to see the complexity decreasing as well as the complexity increasing in the developing language. This is what the present study has demonstrated: there is reduction in complexity and increase in complexity. It may also be important to note that the increase in complexity is an increase in one kind of complexity (expansion of co-emergent lexical and grammatical resources involving greater sophistication of language use) and the reduction in complexity is a reduction in another kind of complexity (a declining number of different patterns of language use).

All this demonstrates the epistemological and methodological value of equally and respectfully considering all instances of language use by the developing language user, regardless of whether or not they are conventional as attested in reference grammars and dictionaries. This is an important point: much of the analysis in current research on 'learner language' either treats innovative instances as part of the developing language user's deficit performance or excludes such occurrences from the analysis, resulting in a partial view of language development. In essence, innovative instances by ELUs constitute part of natural language use, reveal what language resources they have at their disposal at a given point in time as well as how these resources are made use of, and should therefore be given no less attention and consideration in the study of the developing language than the conventional instances of use. This is critical if we are to gain a full appreciation of the complex nature and process of second language development.

5.7 Conclusion

The study has examined a range of patterns of use of the function word *to* in the longitudinal corpus as an attempt to make sense of how second language development takes place over time. It must be noted with caution though that the corpus I work with is relatively small, comprising 496 texts by 124 school students studying English as an additional language. The numbers I am dealing with, particularly in the case of the innovative patterns, are very small, too; so are the changes in frequency observed. Longitudinal data are, however, a major, if not the only, resource through which the process of human language development can be empirically studied. The resulting insights deriving from internal comparisons of samples of the developing language at different points in time are invaluable, especially if the samples are all of a comparable nature or based on the same prompt, as is the case with the longitudinal corpus used in the present study.

Three major findings emerge from the analysis of the longitudinal data in this study. First, there is a proportional increase in the use of inter-clausal *to* constructions over time, in particular the *to*-infinitive clauses. Second, there is a remarkable increase in the use of specific lexical items that are associated with *to*-complement clauses. Together with the clauses, the increase in both quantity and quality of production of the associated lexis indicates far more sophisticated language use observed as time goes by. All this strengthens the findings of the previous study that lexical and grammatical resources grow and construct dynamic interaction with each other, resulting in regularities of association that drive language development over time.

The third and final finding of this study, rather surprisingly, paints a completely different picture of what we know about language development thus far. It concerns the notion of reduction. That is, while the earlier two findings suggest that second language development is about expansion of language resources over time, what has also been observed is that not all language features are found to exhibit an increasing trend of growth. The declining use of six out of eight patterns of use of *to*, with two completely dying out by Time 4, provides empirical evidence to support this view. The finding is based on a consideration which treats all recurring instances of language by ELUs as equally valid patterns of use. This, in turn, leads to the observation of the paradox of complexity in the developing language: In some way, the language is becoming more complex; in some way, it is becoming less complex.

These three findings, in turn, point to two important and interrelated observations about the workings of the developing language as well as the nature and process of second language development. First, two fundamental processes or principles are at work in language development: building up of language resources (a constructive process), and reduction as well as abandonment of some of these resources (a reductionist process). That is, language development is not just about accumulating and storing an ever larger pool of language resources over time; it also involves embracing reduction of some linguistic forms, expressions and patterns as a means of progression. Second, because these two interactive principles of 'give and take' – giving away and taking in of a range of linguistic features – are constantly at work in the course of language development, language use and development by ELUs at all points in time is, arguably, inherently complex. In other words, second language development does not represent a natural progression from simple to more complex language use; after a certain threshold of experience with the relevant language among the developing language users, various points of their development, be they early or later, can be argued to be intrinsically complex throughout the course of language learning, with different points of development signifying complexity in different ways.

These observations converge with, reaffirm and to some extent, go beyond the idea that language development is a nonlinear process (e.g., Larsen-Freeman, 2006, 2009; VanPatten & Williams, 2007; R. Ellis, 2008; Ortega, 2009; N. Ellis, 2011); they may, in fact, provide an explanation for the complex nature and process of second language development. As language development entails a constant two-directional process of construction and reduction, existing acquired structures, expressions and patterns with their corresponding meanings are continuously challenged (i.e., retained, modified or abandoned) and new ones are formed so that at any point in time, the linguistic resources the developing language user enjoys are in a state of flux.

Observations such as these underscore the importance of treating equally all instances of language use by ELUs. If we restrict the description of the developing language to an idealized native-speaker model or any external point of reference, we are unable to uncover and express the two key principles governing complexity in human second language development – what is found in the data is what is chosen to be studied. Once we abandon preconception and witness things as they are, it becomes clear that the two principles are

dynamically at work in the process of language learning and that letting go is no less important than acquiring more resources in the course of language development.

In the next chapter, I present the final corpus study in this thesis and consider what further aspects of the second language developmental process might be revealed through the study of the function word *of*.

Chapter 6

Corpus Study 3: *of*

6.1 Introduction

The previous two studies based on the function words *that* and *to* provide important observations about how ELU language develops over time. It was found, on the one hand, that there was an expansion of the linguistic repertoires over time among the ELUs. This is perhaps unsurprising, or even expected. What is particularly revealing, however, is that the exploration of second language development through high frequency words such as *that* and *to* provides insights into how simultaneous expansion of both lexical and grammatical resources takes place over time. The increased use of *that*-clauses and *to*-clauses in the ELU data, for instance, has been observed to take place in parallel with the use of a wider range of associated lexis. Together they form specific, complex patterns of language identifiable with a diverse and growing range of corresponding meaning groups which have been observed elsewhere in natural language use (Francis at al., 1996, 1998), signifying the expansion of the developing users' language and meaning-making repertoires as time goes by as well as convergence of human language use over time – in this case, language use concerning the English language.

On the other hand, there is, paradoxically, an observed decrease in complexity in the developing language. This seemingly more problematic picture of how language develops emerges in the study of *to*. As shown in Chapter 5, while an apparent linear progression of language development from less to more sophisticated language use is witnessed in the expanding repertoires, there is also an observed declining use of, and a reduction in the number of, different language patterns with which *to* occurs as time goes by. Evidenced in the developing language which is becoming more complex and less complex at the same time in different ways, along different developmental points, second language development can be observed to take place according to two principles: *the construction principle*, which involves acquisition and thus expansion of language resources, and *the reduction principle*, which involves a decline in and/or an elimination of the use of certain linguistic features and patterns.

In this chapter, I consider the use of yet another function word in the developing language observed in the longitudinal data: the wordform *of*. This is the final study in a series of three

attempts using the function words as the lexically driven starting point to understand the second language developmental process. As will be seen, unlike the developmental trends shown by the use of *that* and *to*, there is no consistent change in the use of *of* in the developing language as far as the relative frequency is concerned. As will be shown below, however, consistency in frequency does not necessarily entail consistency in usage.

I begin with a brief review of research relevant to the present study on *to*, before a description follows of the relative frequencies of the word *of* in the developing language as observed in both the longitudinal and cross-sectional corpora. I then present an analysis based on a total of 1,034 concordance lines containing all the instances of use of *of* in LoCDeLUNT, showing how *of* was used by the ELUs over time. Finally, I focus on three major categories of patterns of use of this word in the longitudinal data and consider what the findings reveal about the process of second language development.

6.2 Background

The wordform *of* has been a focus of corpus research for some time. Sinclair (1991), for example, devoted a chapter to the study of this function word in his classic *Corpus, Concordance, Collocation*. He argues that the usage of *of* is not wholly consistent with its usual classification as a preposition: it displays behaviour that is not commonly shared by other prepositions. Rather, *of* might be considered to be the unique exponent of a separate word class. This is part of a general argument Sinclair puts forward concerning the primacy of lexis over grammar. That is, to Sinclair, it is the lexical item that is at the central of language description, not grammatical categories. This view is shared by Hoey, among others, who, in proposing a lexical model or theory of language, suggests that it is the word which is 'primed' to occur with (i.e., 'attracts' the occurrence of) particular other words forming specific collocations, with specific semantic and pragmatic associations, within a specific type of discourse: 'Every word is primed for use in discourse as a result of the cumulative effects of an individual's encounters with the word' (Hoey, 2005, p. 13).

The function word *of* has also been studied in the form of collocational frameworks (e.g., *a. . .of*) in such constructions as *an hour of cooking*, *an agent of evolution* and *an issue of education* in a general corpus (Renouf & Sinclair, 1991) and in a genre-specific corpus (Marco, 2000). The latter study considers the use of such frameworks in a corpus of medical research papers, revealing how some of these discontinuous sequences of words are central to

the phraseology of the discourse studied. The collocates identified for these frameworks have also been found to be motivated by the linguistic conventions of the genre. To use Hoey's (2005) term, the 'priming' effect may have been at work here.

More recently, *of* has been studied in the disciplinary discourses of history and literary criticism by Groom (2007), who finds strong relationships between the phraseology and epistemology of the two disciplines considered through what has been termed semantic sequences (Hunston, 2006). Abstract epistemological constructs have been found to manifest themselves in conventionalized linguistic forms, with grammatical words such as *of*, which are found co-occurring with other elements to form recurrent phraseological units, observed to function as organisers or classifiers of meaning.

The present study considers this high frequency word and extends the analysis to the domain of language acquisition research. I show how patterns of use of this function word change as time goes on and what this suggests about the development of a second language over time.

6.2.1 Frequency of *of* in ELU language

In Chapter 4, we have seen from Table 4.2 that the word *of* occurs as the 13th high frequency word in LoCDeLUNT. How the relative frequency of occurrence of this word changes over time is shown in Table 6.1 below.

Table 6.1: Distribution of *of* across four points of development in the longitudinal corpus

Time 1			Time 2			Time 3			Time 4		
No	**Freq.**	**Per 1000**	**No**	**Freq.**	**Per 1000**	**No**	**Freq.**	**Per 1000**	**No**	**Freq.**	**Per 1000**
13	228	10.18	17	244	8.58	13	304	9.73	16	258	7.89

Unlike *that* (which shows a pattern of increasing use over time) and *to* (which records a consistently high frequency of use over time), the word *of* shows no clear pattern of use over the 24-month period in the ELU data. It is the 13th high frequency word at Time 1, 17th at Time 2, (back to) 13th at Time 3 and later 16th at Time 4. This longitudinal picture of developmental pattern is, however, not observable in the cross-sectional data, as shown in Table 6.2.

Table 6.2: Distribution of *of* across three data sets in the cross-sectional corpus

Primary Five (11-year-olds)			Secondary One (13-year-olds)			Secondary Four (16-year-olds)		
No	**Freq.**	**Per 1000**	**No**	**Freq.**	**Per 1000**	**No**	**Freq.**	**Per 1000**
19	258	7.98	13	525	10.69	9	1131	13.27

As can be seen from the analysis of the cross-sectional data, there is a clear increase in both absolute and relative frequencies in the use of *of* by the younger ELUs to the older ELUs. The word ranks number 19 in the 11-year-old frequency list, 13 in the 13-year-old frequency list and nine in the 16-year-old frequency list. This is an interesting observation, as thus far in the studies of *that* (Chapter 4) and *to* (Chapter 5), the pattern of development of the words as reflected in their respective relative frequency is consistent for both longitudinal and cross-sectional data. The observed difference in the two corpora in this case raises the question of comparability between these two types of data. While some observations are consistent across the data types, others are not. Perhaps a more important issue that arises from the observed difference for SLA and learner corpus researchers, however, is the degree of generalizability and applicability of findings from studies based on a cross-sectional design to inform our understanding of second language development over time. While it is clearly beyond the scope of this study to investigate this issue, it is considered, albeit briefly, in the Future Research section in Chapter 11.

The next section presents an analysis of the various patterns with which *of* occurs and the distribution of each pattern across the four points in time in LoCDeLUNT.

6.3 Analysis of *of*: Patterns and distribution across periods of language development

6.3.1 Patterns of use of *of*

A total of 1034 instances of *of* were observed to occur in the ELU language. 228 of the instances occur at Time 1, 244 at Time 2, 304 at Time 3 and 258 at Time 4. All these instances of *of* were examined to identify patterns of use. Based on a close study of the concordance lines containing all the instances, a total of 14 patterns of use of *of* were identified[2]:

[2] The word 'pattern' is used here irrespective of the number of occurrences.

1. A noun is followed by *of* and a noun phrase: **n *of* n** (e.g., We plucked many *types of flowers*).
2. A verb is followed by *of* and a noun phrase: **v *of* n** (e.g., The girl's parents *praised of Ahmad's kindness*).
3. A verb is followed by a phrasal preposition ending with *of* and a noun phrase: **v phrasal-prep n** (e.g., After a week, Fatin *go out of the hospital*).
4. A verb is followed by a noun and *of* and a noun phrase: **v n *of* n** (e.g., She can't *take care of Muna*).
5. A verb is followed by a noun and a phrasal preposition ending with *of* and a noun phrase: **v n phrasal-prep n** (e.g., Ahmad quickly jump into the river and *pull the girl out of the water*).
6. A verb is followed by *of* and a *that*-clause: **v *of that*-clause** (e.g., While Ah Mei was daydreaming, *dreaming of that she was in a big garden full of flowers*).
7. A verb is followed by *of* and a prepositional *to*-phrase: **v *of to*-phrase** (e.g., Abu ran and *jump of to the river* …).
8. A predicative adjective is followed by *of* and a noun phrase: **pred-adj *of* n** (e.g., They felt very *proud of themselves*).
9. An attributive adjective is followed by *of* and a noun phrase: **attrib-adj *of* n** (e.g., … and have a same hobby which was plucking beautiful and *colourful of flowers* along the river).
10. A phrasal preposition ending with *of* is followed by a noun phrase: **phrasal-prep n** (e.g., She thanked to Danial *because of his kindness*…).
11. A phrasal subordinating conjunction ending with *of* is followed by a clause: **phrasal conj-subord clause** (e.g., *In spite of Muninah was safe*. Atikah was very scared and suddenly fallen down).
12. *of* is followed by a noun phrase: ***of* n** (e.g., …you should called another old man, *of someone* older than you …).
13. Blended patterns: these are sequences of words, some of which morphologically appear to belong to one word class and in terms of pattern, belong to another word class.

    (a) ***a* n/adj *of* n** (e.g., We looked the landscape of the river are so beautiful with *a colourful of flowers*.)

    (b) ***the* n/v *of* n** (e.g., We heard *the shouted of the girl* and saw the incident)

    (c) **adv adj/n *of* n** (e.g., I felt *very pitty of her*)

    (d) **prep n/adj** (e.g., … one of the two girl fall into the river *because of slippery*)

    (e) ***to* n/prep *of* n** (e.g., Then, we helped Ramly to take the girl *to beside of the river*)

(f) **pred-adj *of* n/v** (e.g., She was very *afraid of die* and try to save herself)

(g) **pred-adj *of* n/adj** (e.g., We love this place because it's *full of peaceful*.)

14. fixed phrase (e.g., But *of course* there is something that they both like such as beautiful flowers).

6.3.2 Analyzing and describing innovative patterns of use

The whole process of identifying and describing various patterns of use in the ELU language was not a straightforward matter. As can be seen, many of the 14 patterns of use of *of* were identified and described using the pattern grammar terminology developed by Francis et al (1998) – *many*, but not all. The challenging part comes from the need to account for innovative features of ELU language, which are absent or not attested in the body of language examined by Francis et al (1998) and hence no description and no existing terminology is available, as in other standard reference grammars (e.g., Biber et al., 1999; Carter & McCarthy, 2006). Difficult decisions arise and have to be made concerning how best to go about accounting for their occurrences and introducing new terms to describe the patterns of use with which they occur.

The pattern of a phrasal subordinating conjunction ending with *of* followed by a clause (see Pattern 11 above: **phrasal conj-subord clause**) is a good case in point. The sequence *in spite of* would normally be called a phrasal preposition, but prepositions, by definition, are followed by noun phrases. Thus when we have *in spite of* followed by a clause (as in the example *In spite of Muninah was safe*), the analyst has a difficult decision to make: Is it a phrasal preposition that is followed by a clause, because it is 'in spite of', or do we take the clause to be important, and therefore *in spite of* is no longer a phrasal preposition, but a conjunction, or more precisely, a phrasal subordinating conjunction? There is perhaps no right answer to this question. What I have done is to call it, as just noted, a phrasal subordinating conjunction. That may be said to be more or less an arbitrary decision, but it is also important to note that the classification of instances of use according to relevant patterns has been consistently applied.

The other decision to make is when a feature of language observed in the data appears, morphologically, to belong to one word class and in terms of pattern, to belong to another word class. The category of blended patterns (see, e.g., Pattern 13a above: ***a* n/adj *of* n**) is a good example. The word *colourful* would normally be considered as an adjective, but when it

is presented in the pattern of *a … of* (as in the example *We looked the landscape of the river are so beautiful with a colourful of flowers*), the analyst again has a difficult decision to make: Is it an attributive adjective that is followed by *of* and a noun, because it is 'colourful' (a word ending with *-ful*), or do we give emphasis to the framing pattern of *a … of* so that whichever word is used to 'fill the frame' is considered a noun? What I have decided to do is to call groups of instances of use of this kind blended patterns and to give them both labels of word classes.

It is perhaps also important to point out that there are two instances of use in the ELU data which appear to be candidates for the blended pattern category but which were excluded from such categorization. The first is the word *prouded* in the following sentence:

(1) They *prouded of Ilman* because he safe Farhana life. (097b)

This word is considered an adjective, so is another variant, *pround*, as in example (2) below:

(2) He very *pround of himself* because has saved a life. (049c)

The pattern to which both *prouded* (as in *prouded of Ilman*) and *pround* (as in *pround of himself*) were assigned is **pred-adj *of* n**. They are considered to belong to the same pattern as *proud* in the ELU language:

(3) Their parent *proud of them* for bravery and they done a good job. (053a)
Their family *proud of them* because safe the girl from drowning. (061a)
I *proud of them*. (094b)

The second is the word *scare*, as follows:

(4) … she was also *scare of something* that is more terrible than going alone …. (103c)

Again, this word is considered an adjective and classified as part of the pattern of **pred-adj *of* n**, just as *full*, *rich* and *responsible* are:

(5) Everyday they will go to school together with *full of happiness*. (103b)
The river is very beautiful and *rich of fish*. (049c)
"… villagers are so *responsible of their river, wild and animals*," Oprah replied. (051d)

Table 6.3: Distribution of patterns of use of *of* in the longitudinal corpus

No	Pattern	Time 1		Time 2		Time 3		Time 4	
1	**n *of* n**	188	(%) 82.5	211	(%) 86.5	246	(%) 80.9	220	(%) 85.3
2	**v *of* n**	4	1.8	4	1.6	5	1.6	2	0.8
3	**v phrasal-prep n**	2	0.9	1	0.4	2	0.7	3	1.2
4	**v n *of* n**	2	0.9	-	-	5	1.6	3	1.2
5	**v n phrasal-prep n**	1	0.4	1	0.4	4	1.3	1	0.4
6	**v *of that*-clause**	1	0.4	-	-	-	-	-	-
7	**v *of to*-phrase**	1	0.4	-	-	-	-	-	-
8	**pred-adj *of* n**	14	6.1	15	6.1	17	5.6	9	3.5
9	**attrib-adj *of* n**	1	0.4	1	0.4	2	0.7	-	-
10	**phrasal prep n**	13	5.7	7	2.9	19	6.3	16	6.2
11	**phrasal conj-subord clause**	-	-	1	0.4	2	0.7	1	0.4
12	***of* n**	-	-	1	0.4	-	-	-	-
13	**Blended pattern** **(a) *a* n/adj *of* n** **(b) *the* n/v *of* n** **(c) adv adj/n *of* n** **(d) prep n/adj** **(e) *to* n/prep *of* n** **(f) pred-adj *of* n/v** **(g) pred-adj *of* n/adj**			1		1		1 1	
	Sum	1	0.4	2	0.8	2	0.7	2	0.8
14	**Fixed phrase**	-	-	-	-	-	-	1	0.4
	Total	**228**	**100**	**244**	**100**	**304**	**100**	**258**	**100**

6.3.3 Distribution of patterns

When all the decisions described in Section 6.3.2 had been made, the instances of each of these 14 patterns in each time period of the longitudinal corpus were then counted. These results are presented in Table 6.3.

As can be seen, the dominant pattern observed in the ELU language throughout the four time points of development is the **n *of* n** pattern. It accounts for 82.5% of all the *of* patterns at Time 1, 86.5% at Time 2, 80.9% at Time 3 and 85.3% at Time 4. There are no consistent changes in use observed as far as **n *of* n** is concerned, except to note that the ELUs were found to make a slight greater use of noun phrases at Time 4 (85.3%) than when they were at Time 1 (82.5%). This is perhaps not very surprising because noun phrases can be considered to be basic language resources ELUs draw upon to express general ideas. Similar results were observed of a dominant use of **n *of* n** among patterns of use of the wordform *of* in the study of a general corpus (Sinclair, 1991) and in a discipline-specific corpus (Groom, 2007).

The next three categories of patterns of use of *of*, after **n *of* n**, are patterns associated with verbs, adjectives and phrasal prepositions. The patterns with verbs (hereafter the 'verb patterns') are those shown as Patterns 2-7 in Table 6.3 while the patterns with adjectives (hereafter the 'adjective patterns') are those shown as Patterns 8-9 in the same table. For discussion purposes, they are classified as two main categories and presented in Table 6.4 below, reproduced from the results in Table 6.3, together with the pattern of phrasal preposition followed by a noun phrase (Pattern 10 in Table 6.3).

As can be seen, there are no clear trends in the proportional use of these patterns over time. The verb patterns, for example, occur 4.8% at Time 1, 2.4% at Time 2, 5.2% at Time 3 and 3.6% at Time 4. The adjective patterns occur with a rather consistent proportion of use at Times 1-3 but a decrease in proportion is observed at Time 4. The **phrasal-prep n** pattern, on the other hand, shows a decrease in proportion at Time 2 but a similar proportion of use at the other time periods. When the results at Time 1 are compared with those at Time 4, however, we may say that there is a slight decline in use for the verb patterns (4.8% at Time 1 versus 3.6% at Time 4), a sharp decline in use for the adjective patterns (6.5% at Time 1 versus 3.5% at Time 4), and a slight increase in use for **phrasal-prep n** (5.7% at Time 1 versus 6.2% at Time 4).

Table 6.4: Distribution of categories of patterns of use of *of* with verbs, adjectives and phrasal prepositions

Pattern	**Time 1**		**Time 2**		**Time 3**		**Time 4**	
Patterns associated with verbs (2) v *of* n (3) v phrasal-prep n (4) v n *of* n (5) v n phrasal-prep n (6) v *of that*-clause (7) v *of to*-phrase	11	% 4.8	6	% 2.4	16	% 5.2	9	% 3.6
Patterns associated with adjectives (8) pred-adj *of* n (9) attrib-adj *of* n	15	6.5	16	6.5	20	6.6	9	3.5
phrasal-prep n	13	5.7	7	2.9	19	6.3	16	6.2

The next category of observed patterns, occurring with the least proportion of use, comprises **phrasal conj-subord clause**, ***of* n**, blended patterns and the fixed phrase *of course*. Their frequency and proportion of use are presented in Table 6.5, again reproduced from the results in Table 6.3.

Table 6.5: Distribution of low frequency patterns of *of*

Pattern	**Time 1**		**Time 2**		**Time 3**		**Time 4**	
phrasal conj-subord clause	-	% -	1	% 0.4	2	% 0.7	1	% 0.4
***of* n**	-	-	1	0.4	-	-	-	-
Blended pattern	1	0.4	2	0.8	1	0.3	2	0.8
Fixed phrase	-	-	-	-	-	-	1	0.4

Together these low frequency patterns account for less than 2.0% of all the patterns of use of *of* at each point in time over the 24-month period of language development among the ELUs.

Except for the fixed phrase, the first three patterns are innovative features in the ELU language.

6.4 Observing language development through ELUs' uses of *of* over time

The patterns of use of *of* have thus far been approached from a quantitative perspective. This section takes a closer look at instances of use of three major categories of patterns of *of* and explores what they reveal about the process of second language development. The categories are:

1. n *of* n
2. The verb patterns
3. The adjective patterns

6.4.1 The n *of* n pattern

In studying the **n *of* n** pattern, I am interested in examining changes in the use of nouns and noun phrases over time with this particular pattern. As in the previous two studies reported in Chapters 4 and 5, only those instances that occur at least twice and in two different ELU texts are considered. Of the 865 instances of the **n *of* n** pattern observed in LoCDeLUNT and studied in this chapter, 522 were found to occur in the form of what Renouf and Sinclair (1991) have termed collocational frameworks and meet the criteria of frequency and range established in this thesis.

Renouf and Sinclair define collocational frameworks as follows:

> … 'frameworks' consist of a discontinuous sequence of two words, positioned at one word remove from each other; they are therefore not grammatically self-standing; their well-formedness is dependent on what intervenes. (Renouf & Sinclair, 1991, p. 128)

The examples of the frameworks studied by Renouf and Sinclair (1991) are:

a + ? + *of*	*be* + ? + *to*	*for* + ? + *of*
an + ? + *of*	*too* + ? + *to*	*had* + ? + *of*
		many + ? + *of*

As the focus of this sub-section is on **n *of* n**, it is perhaps not surprising that *a* + ? + *of* and *many* + ? + *of* were also observed to occur in the ELU data. Examples from LoCDeLUNT include *a bouquet of flowers* and *many types of flowers*, respectively. The framework of *an* + ?

+ *of* is, however, not attested in the data; what was observed in LoCDeLUNT instead is the framework *the* + ? + *of*, an example of which is *the side of the river*.

Of the 522 instances of recurring collocational frameworks, 120 occur at Time 1, 134 at Time 2, 130 at Time 3 and 138 at Time 4. They have been categorized into five groups, as follows:

1. quantity + *of* + n (e.g., *one of the (three) girls*)
2. *the* + quantity + *of* + n (e.g., *the three/two of them*)
3. *many* + category + *of* + n (e.g., *many types of flowers*)
4. *a*(*nother*)/*the* + n + *of* + n (e.g., *a bouquet of flowers, the shore of the lake*)
5. n + *of* + n (e.g., side of the lake)

It must be pointed out that the categorization of the frameworks into five groups is not a straightforward matter. Challenges arise pertaining to reducing a large amount of disparate data to a few groups. The classification is, to a large extent, arbitrary. As can be seen, the categorization has been based on a focus on the first noun (e.g., whether or not there is a determiner preceding it, as found in the data) and not on the second noun phrase. All the observed instances of use of the frameworks have, however, been identified and grouped consistently according to the features of the first noun they share. Frameworks such as *both of them* and *all of them/us* are, therefore, occurrences considered in the quantity + *of* + n group while frameworks such as *the three/two of them* and *the three of boys* in the *the* + quantity + *of* + n group.

Among these groups of frameworks, Groups 1 – 4 may reflect conventional language use while Group 5 innovative language use. The results of the analysis and classification are presented in Table 6.6.

Table 6.6: Distribution of the instances of n *of* n pattern across four points of language development

Time 1		Time 2		Time 3		Time 4	
	Freq		Freq		Freq		Freq
quantity + *of* + n		**quantity + *of* + n**		**quantity + *of* + n**		**quantity + *of* + n**	
one of the (three) girls/boys	39	one of the (three) girls/boys	37	one of the (three) girls/boys	31	one of the (three) girls/boys	32
one of them	25	one of the girl/boy	16	all of them/us	13	one of them	17
one of the (two) boy/girl	17	one of them	15	one of the girl/boy	11	one of the girl/boy	14
both of them	4	all of them	5	one of them	9	both of you/them/us	8
a lot of flowers	2	both of them	5	both of them/us/you	8	all of them/you	5
a lot of money/work	2	a lot of fish	4	a lot of water	2	some of them	3
both of the girls	2	both of the girls	4	both of the girls	2	all of the stuffs/equipments	2
one of the men/women	2	a lot of flowers	2	each of them	2	some of the flower	2
		a few of flowers	2	many of flower	2	some of the flowers	2
		a lot of water/rest	2	some of them	2		
		one of the women/men	2	three of them/us	2		
				***the* + quantity + *of* + n**		***the* + quantity + *of* + n**	
				the three/two of them	5	the three/two of them	9
				the three of boys	2		
***many* + category + *of* + n**		***many* + category + *of* + n**		***many* + category + *of* + n**		***many* + category + *of* + n**	
many types of flower	7	many types of flowers	6	many types of flowers	5	many types of flowers	7
many kind of flowers	3	many type of flowers	3				
many types of flowers	3	many types of flower	2				
***the* + n + *of* + n**		***a/the* + n + *of* + n**		***a*(*nother*)/*the* + n + *of* + n**		***a/the* + n + *of* + n**	
the shore of the lake/river	4	a bouquet of flower	3	a bouquet of flowers/roses	3	a bouquet of flowers/tulips	3
the bank of river	2	a bouquet of flowers	2	a bouquet of flower	2	a group of boys/friends	3
the bank of the river	2	a pail of baits/worms	2	a cup of coffee	2	a pail of (tape)worms	3
the moral values of this story	2	the side of the lake/river/pond	8	another location of the lake	2	a bunch of flowers	2
the water of the lake/river	2	the bank of the river/lake	5	a pail of worms	2	a pair of trousers/eyes	2
		the water of the river	5	a twinkling of an eye	2	a school of fish	2
n + *of* + n		the shore of the river/pond/lake	4	the side of the river/lake/pond	10	the side of the river/lake	7
side of the lake/river	2			the edge of the river	3	the bank of the river/miningpool	6
				the shore of the lake/river	3	the shore of the river/lake	4
				the bank of the river	3	the surface of the water/lake/river	3
				the band of the river/lake	2	the parents of the girl	2
17	120	21	134	25	130	22	138

One observation from Table 6.6 is that the number of groups of collocational frameworks, observed along the various developmental points in this study, starts with four at Time 1, changes to three at Time 2, before increasing to four groups again at Time 3 and Time 4. While there are four groups of frameworks each observed at Time 1, Time 3 and Time 4, only three groups are common across these three points of development: (1) quantity + *of* + n (e.g., *one of the (three) girls/boys*), (3) *many* + category + *of* + n (e.g., *many types of flowers*), and (4) *a*(*nother*)/*the* + n + *of* + n (e.g., *a bouquet of flowers*). The fourth, specific group observed only at Time 1 is the innovative n + *of* + n group (e.g., side of the lake/river). This innovative group is not observable at the next point of development, Time 2, which as noted above, witnesses only the production of three groups of frameworks. This specific group also does not make a re-appearance at the next two later points of development, Time 3 and 4, whose shared, distinctive group is the conventional *the* + quantity + *of* + n group (e.g., *the three/two of them*). There seems to be a change then over time from a relatively more innovative language use at Time 1 to a more conventional language use by Time 4 in the production of these frameworks.

There is also an increase in the number of types of frameworks used as time goes on, from 17 at Time 1 to 22 at Time 4, although the greatest number of types is found at Time 3 (i.e., 25 in total). More qualitative differences can be observed when each of the five categories of use is examined across the four time periods of development.

1. quantity + *of* + n

Table 6.7 shows the use of instances of the quantity + *of* + n framework by the ELUs in the study. The results can be considered from the perspectives of (a) the use of quantifiers, (b) the use of noun phrases and (c) the way the ELUs use (a) and (b) together.

Table 6.7: Instances of use in the quantity + *of* + n framework

Time 1	Time 2	Time 3	Time 4
1. one of the (three) girls/boys	1. one of the (three) girls/boys	1. one of the (three) girls/boys	1. one of the (three) girls/boys
2. one of them	2. one of the girl/boy	2. all of them/us	2. one of them
3. one of the (two) boy/girl	3. one of them	3. one of the girl/boy	3. one of the girl/boy
4. both of them	4. all of them	4. one of them	4. both of us/you/them
5. a lot of flowers	5. both of them	5. both of us/you/them	5. all of them/you
6. a lot of money/work	6. a lot of fish	6. a lot of water	6. some of them
7. both of the girls	7. both of the girls	7. both of the girls	7. all of the stuffs/equipments
8. one of the men/women	8. a lot of flowers	8. each of them	8. some of the flower
	9. a few of flowers	9. many of flower	9. some of the flowers
	10. a lot of water/rest	10. some of them	
	11. one of the women/men	11. three of them/us	

(a) The use of quantifiers

As far as quantifiers are concerned, *all, both* and *one* occur throughout the four time points of development (see Table 6.7a). This suggests that they have been emergent properties in the developing language since Time 1. The quantifier *a lot*, however, is characteristic only in the ELU language at Time Points 1-3; it does not feature in the language at Time 4. Along the points of development, an increasing number of quantifiers make their appearances, with the greatest range of quantifiers being observed at Time 3 (*all, a lot, both, one, some, many, three*). The ELUs, however, show a more selective use of quantifiers when they reach Time 4.

Table 6.7a: The use of quantifiers in the quantity + *of* + n framework

Time 1	Time 2	Time 3	Time 4
? + *of* + n	? + *of* + n	? + *of* + n	? + *of* + n
1. all 2. a lot 3. both 4. one	1. all 2. a few 3. a lot 4. both 5. one	1. all 2. a lot 3. both 4. one 5. some 6. many 7. three	1. all 2. both 3. one 4. some

(b) The use of noun phrases

The results obtained when the noun phrases are the focus of study reveal potentially interesting insights into second language development. As can be seen from Table 6.7b, the third-person pronoun *them* dominates earlier language use at Time 1 and Time 2, but a richer use of the pronoun system with *of*, including the employment of the first-person pronoun *us* and second-person pronoun *you*, becomes evident at later points of development (Time 3 and Time 4). This emergent use of *you* at a later point in time is particularly noteworthy, as this may suggest how the ELUs were changing, reconceptualizing and reconstructing their ways of presenting their narratives over time, likely from an initial tendency to present a 'monologic' third-person narrative account (i.e., through the use of *them*) to one which is increasingly more 'dialogic', involving illustrations of conversations between the main characters in the narrative (i.e., through the use of *you*). This is, however, not supported in the findings of overall frequency of *you* observed at different points in time in the longitudinal data: there is only a very slight increase in the use of *you a*nd *your*, from 0.11% at Time 1 to 0.12% at Time 4.

Table 6.7b: The use of noun phrases in the quantity + *of* + n framework

Time 1	Time 2	Time 3	Time 4
quantity + *of* + ?	quantity + *of* + ?	quantity + *of* + ?	quantity + *of* + ?
1. the (three) girls/boys	1. the (three) girls/boys	1. the (three) girls/boys	1. the (three) girls/boys
2. them	2. the girl/boy	2. them	2. them
3. the (two) boy/girl	3. them	3. the girl/boy	3. the girl/boy
4. flowers	4. fish	4. us	4. us
5. money	5. flowers	5. you	5. you
6. work	6. water	6. water	6. stuffs
7. the men/women	7. rest	7. flower	7. equipments
	8. the women/men		8. flower
			9. flowers

In addition to the use of pronouns and such nouns as *boys, girls, men* and *women*, the range of noun phrases employed includes *flowers, money* and *work* (Time 1), *fish, flowers, water* and *rest* (Time 2), *water* and *flower* (Time 3) and *stuffs, equipments, flower* and *flowers* (Time 4). Differences between Time 1 and Time 4 are small, suggesting that this is not an area where a large expansion of repertoire is evident, but where a small amount of expansion can be observed.

It is also important to note that there are only two noun phrases (*water* and *flower*) observed at Time 3. This is surprising because while the early analysis of the use of quantifiers suggests that the ELUs were most productive at Time 3 (a point in time where they used the widest range of quantifiers), the results here show otherwise: they used the smallest number of noun phrases at this same point among all the four periods of development. This perhaps reinforces an important lesson noted in the previous chapter about the whole research enterprise: what is found is dependent on what the analyst chooses to quantify.

(c) The holistic use of the framework

When both the quantifiers and noun phrases are considered together, a different picture of language development emerges. Most productive language use, as measured in terms of the number of different sequences produced, is witnessed at Time 2 and Time 3. A total of 11 different sequences emerge at each of these points of development, as shown in Table 6.7c.

Table 6.7c: The holistic use of the quantity + *of* + n framework

Time 1	Time 2	Time 3	Time 4
1. one of the (three) girls/boys	1. one of the (three) girls/boys	1. one of the (three) girls/boys	1. one of the (three) girls/boys
2. one of them	2. one of the girl/boy	2. all of them/us	2. one of them
3. one of the (two) boy/girl	3. one of them	3. one of the girl/boy	3. one of the girl/boy
4. both of them	4. all of them	4. one of them	4. both of us/you/them
5. a lot of flowers	5. both of them	5. both of us/you/them	5. all of them/you
6. a lot of money/work	6. a lot of fish	6. a lot of water	6. some of them
7. both of the girls	7. both of the girls	7. both of the girls	7. all of the stuffs/equipments
8. one of the men/women	8. a lot of flowers	8. each of them	8. some of the flower
	9. a few of flowers	9. many of flower	9. some of the flowers
	10. a lot of water/rest	10. some of them	
	11. one of the women/men	11. three of them/us	

There are two observations to note here. First, there is a shift in lexical choices over time for the collocational frameworks involving *flower* and *flowers*. While *a lot of flowers* and *a few of flowers* feature in the language at Time 1 and Time 2, *many of flower, some of the flower* and *some of the flowers* are observed at Time 3 and Time 4. Second, there is a small, gradual increase in the use of alternative forms of expressions over time in the ELU language. Both *one of the girls* and *one of the girl*, for example, occur at Time 1; they are also found to recur throughout all the later points of development. At Time 4, there is, additionally, another set of expressions, *some of the flowers* and *some of the flower*, observed in the language. Overall, however, the ELUs are found to be most productive in language use at Time 2 and Time 3, as noted above.

2. *the* + quantity + *of* + n

Table 6.8 shows the use of instances of the *the* +quantity + *of* + n framework by the ELUs in the study. We find that the framework started to emerge only at Time 3, with *the three/two of them* and *the three of boys* being the only two expressions used at this point in time. At Time 4, however, only the expression of *the three/two of them* remains in the ELU language.

Table 6.8: Instances of use in the *the* +quantity + *of* + n framework

Time 1	Time 2	Time 3	Time 4
		1. the three/two of them	1. the three/two of them
		2. the three of boys	

3. *many* + category + *of* + n

Table 6.9 shows the use of instances of the *many* + category + *of* + n framework by the ELUs in the study. As can be seen, there is a gradual reduction over time in the number of different sequences produced in the ELU data, from three each at Time 1 and Time 2 to one each at Time 3 and Time 4. All the sequences shown are different in form but entail the same proposition: they are all concerned with the idea of 'variety of flowers'. A range of alternative expressions were observed at Time 1 and Time 2, including *many types of flower, many kind of flowers, many type of flowers* and *many types of flowers*. At later points of development (i.e., Time 3 and Time 4), however, only the expression *many types of flowers* remains in the ELU language. While the findings of the analysis of the earlier two frameworks are indicative of an expansion of linguistic repertoire taking place over time, here we find evidence of a reduction process at work in second language development (see Chapter 5 for more discussion).

Table 6.9: Instances of use in the *many* + category + *of* + n framework

Time 1	Time 2	Time 3	Time 4
1. many types of flower 2. many kind of flowers 3. many types of flowers	1. many types of flowers 2. many type of flowers 3. many types of flower	1. many types of flowers	1. many types of flowers

4. *a*(*nother*)/*the* + n + *of* + n

Table 6.10 shows the use of instances of the *a*(*nother*)/*the* + n + *of* + n framework by the ELUs in the study. The results can be considered from the perspectives of (a) the use of nouns, (b) the use of noun phrases and (c) the way the ELUs use (a) and (b) together.

Table 6.10: Instances of use in the *a*(*nother*)/*the* + n + *of* + n framework

Time 1	Time 2	Time 3	Time 4
1. the shore of the lake/river 2. the bank of river 3. the bank of the river 4. the moral values of this story 5. the water of the lake/river	1. a bouquet of flower 2. a bouquet of flowers 3. a pail of baits/worms 4. the side of the lake/river/pond 5. the bank of the river/lake 6. the water of the river 7. the shore of the river/pond/lake	1. a bouquet of flowers/roses 2. a bouquet of flower 3. a cup of coffee 4. another location of the lake 5. a pail of worms 6. a twinkling of an eye 7. the side of the river/lake/pond 8. the edge of the river 9. the shore of the lake/river 10. the bank of the river 11. the band of the river/lake	1. a bouquet of flowers/tulips 2. a group of boys/friends 3. a pail of (tape)worms 4. a bunch of flowers 5. a pair of trousers/eyes 6. a school of fish 7. the side of the river/lake 8. the bank of the river/miningpool 9. the shore of the river/lake 10. the surface of the water/lake/river 11. the parents of the girl

(a) The use of nouns in the first slot of n *of* n

As shown in Table 6.10a, there is a steady increase in the number of noun-types used over time among the ELUs. At Time 1, for instance, only four different nouns are observed in the data. By Time 4, this number has increased over two times the initial number of different nouns, which were produced 24 months earlier. The greater number of different nouns found over time suggests an expanding repertoire among the users to convey ideas using the language resources at hand. It also seems to be the case that the developing users are moving from using relatively fixed phrases (*the bank of the river*) to a productive exploitation of the framework. Arguably, they have started to see a framework into which nouns can be slotted, rather than a few fixed phrases.

Table 6.10a: The use of first slot of nouns in the *a*(*nother*)/*the* + n + *of* + n framework

Time 1	**Time 2**	**Time 3**	**Time 4**
a/the + ? + *of* + n	a/the + ? + *of* + n	a(nother)/the + ? + *of* + n	a/the + ? + *of* + n
1. shore	1. bouquet	1. bouquet	1. bouquet
2. bank	2. pail	2. cup	2. group
3. water	3. side	3. location	3. pail
4. moral values	4. bank	4. pail	4. bunch
	5. water	5. twinkling	5. pair
	6. shore	6. side	6. school
		7. bank	7. side
		8. shore	8. bank
		9. edge	9. shore
		10. band	10. surface
			11. parents

(b) The use of noun phrases in the second slot of n *of* n

When the use of noun phrases is examined from the longitudinal data, we find again a clear incremental pattern of development. As shown in Table 6.10b, the ELUs began with the use of four different noun phrases at Time 1, increased the use of different noun phrases to six and nine at Time 2 and Time 3, respectively, before producing 13 different noun phrases at Time 4. Also, more specific noun phrases such as *roses* and *tulips*, which constitute hyponyms of *flowers* used from Time 2 onwards, are observed to emerge in the ELU language at later points of development (i.e., at Time 3 and Time 4).

Table 6.10b: The use of second slot of noun phrases in the *a*(*nother*)/*the* + n + *of* + n framework

Time 1	Time 2	Time 3	Time 4
a/the + n + *of* + ?	a/the + n + *of* + ?	A(nother)/the + n + *of* + ?	a/the + n + *of* + ?
1. the lake	1. flower	1. flowers	1. flowers
2. river	2. flowers	2. roses	2. tulips
3. the river	3. baits	3. flower	3. boys
4. this story	4. worms	4. coffee	4. friends
	4. the lake	5. the lake	5. worms
	5. the river	6. worms	6. trousers
	6. the pond	7. an eye	7. eyes
		8. the river	8. fish
		9. the pond	9. the river
			10. the lake
			11. the miningpool
			12. the water
			13. the girl

(c) The holistic use of the framework

When both the nouns and noun phrases are considered together, what is obvious is that an expansion of language resources can be seen to have taken place over time. In addition, it can be argued that a reinterpretation of the resources has been observed in which a productive exploitation of the framework is observed, from the initial few, relatively fixed phrases to the later more flexibly produced sequences. As shown in Table 6.10c, most productive language use, as measured in terms of the number of different sequences produced, is witnessed at Time 3 and Time 4. A total of 11 different sequences emerge at each of these two later points of development.

Table 6.10c: The holistic use of the *a*(*nother*)/*the* + n + *of* + n framework

Time 1	Time 2	Time 3	Time 4
1. the shore of the lake/river	1. a bouquet of flower	1. a bouquet of flowers/roses	1. a bouquet of flowers/tulips
2. the bank of river	2. a bouquet of flowers	2. a bouquet of flower	2. a group of boys/friends
3. the bank of the river	3. a pail of baits/worms	3. a cup of coffee	3. a pail of (tape)worms
4. the moral values of this story	4. the side of the lake/river/pond	4. another location of the lake	4. a bunch of flowers
5. the water of the lake/river	5. the bank of the river/lake	5. a pail of worms	5. a pair of trousers/eyes
	6. the water of the river	6. a twinkling of an eye	6. a school of fish
	7. the shore of the river/pond/lake	7. the side of the river/lake/pond	7. the side of the river/lake
		8. the edge of the river	8. the bank of the river/miningpool
		9. the shore of the lake/river	9. the shore of the river/lake
		10. the bank of the river	10. the surface of the water/lake/river
		11. the band of the river/lake	11. the parents of the girl

5. n + *of* + n

Table 6.11 shows the use of instances of the n + *of* + n framework by the ELUs in the study. There is only one instance of use witnessed of this 'zero-article' framework, *side of the lake/river*, and it is found at Time 1.

Table 6.11: Instances of use in the n + *of* + n framework

Time 1	Time 2	Time 3	Time 4
1. side of the lake/river			

6.4.2 The verb patterns

We now turn to the use of verb patterns with *of* by the ELUs in this study. As noted above, there are six relevant patterns observed in the data. They are:

1. **v *of* n** (e.g., The girl's parents *praised of Ahmad's kindness*).
2. **v phrasal-prep n** (e.g., After a week, Fatin *go out of the hospital*).
3. **v n *of* n** (e.g., She can't *take care of Muna*).
4. **v n phrasal-prep n** (e.g., Ahmad quickly … *pull the girl out of the water*).
5. **v *of that*-clause** (e.g., While Ah Mei was daydreaming, *dreaming of that she was in a big garden full of flowers*).
6. **v *of to*-phrase** (e.g., Abu ran and *jump of to the river* …).

How changes in the ELUs' production of each of these verb patterns take place over time is considered below.

1. v *of* n

Table 6.12 shows the use of instances of the v + *of* + n pattern by the ELUs in the study. As can be seen, this is a pattern of use most productively applied at Time 3 and surprisingly least so at Time 4. Verbs observed at each point in time are *felt, saved* and *praised* at Time 1, *depend, praise* and *thinking* at Time 2, *caught, heard, knew, pluck* and *strolling* at Time 3, and *thinking* and *told* at Time 4. What is puzzling here is that there seems to be little continuity in the verbs used. *THINK of*, for instance, occurs at Time 2 and Time 4, but not at Time 3. *TELL of*, on the other hand, occurs only at Time 4.

Over time, there seems to be a shift towards richer use of noun phrases in the language in relation to this verb pattern. While concrete nouns such as *the shore* and abstract nouns such as *kindness* are observed across the various points of development, the emergence of the *–ing* clauses such as *Lina's shouting* at Time 3 (as in *heard of Lina's shouting*) and *giving those flowers* at Time 4 (as in *thinking of giving those flowers to their teacher*), suggests an expanded repertoire of noun phrases that go beyond just common nouns.

Table 6.12: The use of v + *of* + n in the ELU data

Time 1	Time 2	Time 3	Time 4
1. felt of the shore 2. saved of him 3. praised of Ahmad's kindness 4. praised of her brave	1. depend of you 2. praise of Aizat 3. praise of Aizat brave and kind heart 4. thinking of himself	1. caught of her 2. heard of Lina's shouting 3. knew quite well of her mother taste 4. pluck of flower 5. strolling of the park	1. thinking of giving those flowers to their teacher 2. told of Ali's generosity

2. v phrasal-prep n

Table 6.13 shows the use of instances of the v phrasal-prep n pattern by the ELUs in the study. There is a small increase in the use of this verb pattern over time but the pattern is, as can be seen, not a particularly productive one.

Table 6.13: The use of v phrasal-prep n in the ELU data

Time 1	Time 2	Time 3	Time 4
1. go out of the hospital 2. walking back of them	1. ran out of their class	1. checked up of anything 2. screamed out of voices	1. moved out of the hole 2. pulled out of the water 3. screamed out of the voice

3. v n *of* n

Table 6.14 shows the use of instances of the v n *of* n pattern by the ELUs in the study. As can be seen, the pattern is based on the collocation 'take care'. There is no instance of use observed at Time 2; in fact, there is no clear developmental pattern here, except to note that there is an apparent shift in the use of noun phrases from a third-person perspective (e.g., *her sister* and *theirself*) at Time 1 to encompassing a first- and second-person perspective (e.g., *myself* and *yourselves*) at Time 3.

Table 6.14: The use of v n *of* n in the ELU data

Time 1	Time 2	Time 3	Time 4
1. take good care of her sister 2. took care of theirself	-	1. take care of herself (x2) 2. take care of Muna 3. take care of myself 4. take care of yourselves	1. take care of herself 2. taking care of my own neighbour 3. take care of themselves

4. v n phrasal-prep n

Table 6.15 shows the use of instances of the v n phrasal-prep n pattern by the ELUs in the study. This pattern is based on a small number of verbs (*PULL* and *TAKE*) relating to the idea of getting someone out of the water, and is most productively used at Time 3.

Table 6.15: The use of v n phrasal-prep n in the ELU data

Time 1	Time 2	Time 3	Time 4
1. pull the girl out of the water	1. took Siti out of the river	1. pull the girl out of the river 2. pulled Mira out of the water 3. pulled the girl out of the lake 4. take the girl out of the river	1. pull Aini out of the river

5. v *of that*-clause

Table 6.16 shows the use of the v *of that*-clause pattern observed in the study. Only one instance of use is found and it occurs at Time 1.

Table 6.16: The use of v *of that*-clause in the ELU data

Time 1	Time 2	Time 3	Time 4
1. dreaming of that she was in a big garden full of flowers	-	-	-

6. v *of to*-phrase

Table 6.17 shows the use of the v *of to*-phrase pattern observed in the study. Like the previous verb pattern, only one instance of use of this pattern is found and it occurs at Time 1.

Table 6.17: The use of v *of to*-phrase in the ELU data

Time 1	Time 2	Time 3	Time 4
1. jump of to the river	-	-	-

6.4.3 The adjective patterns

We now consider the use of adjective patterns with *of* by the ELUs in this study. Two patterns have been observed in the data. They are:

1. **pred-adj *of* n** (e.g., They felt very *proud of themselves*).
2. **attrib-adj *of* n** (e.g., … was plucking beautiful and *colourful of flowers* along the river).

How changes in the ELUs' production of each of these patterns take place over time is considered below.

1. pred-adj *of* n

Table 6.18 shows the use of the pred-adj *of* n pattern by the ELUs in the study. Most productive language use is observed at Time 3 (17 tokens, 12 types), followed by that observed at Time 2 (15 tokens, 12 types). In terms of the number of different adjectives used, Time 3 also indicates a point in development where the greatest number of adjectives is produced, including *full, proud, rich* and *scare*.

Table 6.18: The use of pred-adj *of* n in the ELU data

Time 1	Time 2	Time 3	Time 4
1. full of flowers 2. proud of him (x3) 3. proud of himself (x2) 4. proud of the three boys quick action 5. proud of them (x4) 6. proud of themselves (x2) 7. proud of us	1. full of happiness 2. full of variety types of fishes 3. full of will 4. proud of their act 5. proud of their bravery 6. proud of their son 7. proud of their sons 8. proud of them (x4) 9. proud of themselves 10. proud of you 11. proud of Syaz 12. prouded of Ilman	1. full of fish 2. full of fishes 3. full of worms 4. proud of Daniel's self-sacrifice 5. proud of him (x4) 6. proud of himself (x2) 7. proud of me 8. proud of their son 9. proud of them (x2) 10. pround of himself 11. rich of fish 12. scare of something	1. full of baits and three old rods 2. full of fishes and lilies 3. proud of him 4. proud of their bravery 5. proud of them (x2) 6. proud of themselves 7. proud of you 8. responsible of their river

An interesting observation here concerns the difference in language performance between Time 1 and Time 4. Language use at these two points in time appears to be similar at first sight but a closer look indicates otherwise. While there are 14 tokens of sequences or 7 types of expressions observed at Time 1, there are nine tokens or eight types of expressions found at Time 4. In other words, the ELUs began with a small number of expressions which were used relatively frequently and changed, after a 24-month period, to using a similar number of expressions but at this point in time, rather infrequently. Qualitatively there is also a sense of greater specificity in the use of noun phrases at this later point in time, including the use of

baits and three old rods and *fishes and lilies*. In between the time frame, as noted above, there is a flood of expressions observed at Time 2 and Time 3.

2. attrib-adj *of* n

Table 6.19 shows the use of the attrib-adj *of* n pattern by the ELUs in the study. This adjective pattern is, as can be seen, an area not particularly productive; in fact, it dies out by Time 4.

Table 6.19: The use of attrib-adj *of* n in the ELU data

Time 1	Time 2	Time 3	Time 4
1. assorted of flowers	1. colourful of flowers	1. colourful of flower 2. colourful of flowers	

6.5 Understanding second language development: Fundamental patterns of growth

The picture of language development emerging from the study of the function word *of* in the longitudinal data, with its three major categories of patterns of use considered in detail, is a complex one. It is complex because there is no single dominant pattern of development that is observable over time. What the findings of this study suggest, based on an analysis of all the 1,034 concordance lines of *of* in the ELU data, is that second language development may, depending on the language features in focus, be characterized by four key developmental patterns over time. They are:

1. a three-phase pattern
2. an expansion pattern
3. a reduction pattern
4. no consistent pattern

1. A three-phase pattern

A three-phase pattern is observed when a language feature is found to begin with few expressions and then proceed to a flood of expressions, before a downward trend in use of the features takes place. Diagrammatically this developmental pattern may be represented as follows:

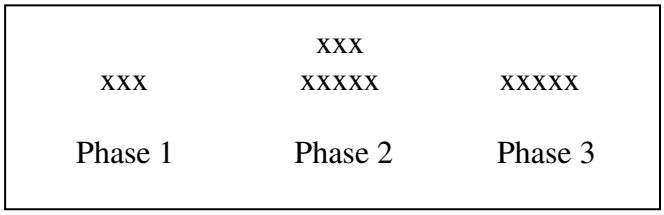


Figure 6.1 A three-phase pattern in second language development

In this study, the quantity + *of* + n framework (e.g., *some of them*) and the pred-adj *of* n pattern (e.g., *proud of their bravery*) are two language features exemplifying this developmental pattern.

2. An expansion pattern

An expansion pattern is observed when a language feature is found to 'grow' or develop from few to more expressions over time. Diagrammatically this developmental pattern may be represented as follows:

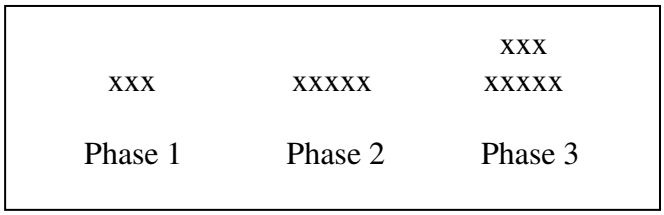


Figure 6.2 An expansion pattern in second language development

In this study, the *a*(*nother*)/*the* + n + *of* + n framework (e.g., *a bouquet of flowers*) is a feature exemplifying this developmental pattern.

3. A reduction pattern

A reduction pattern is observed when a language feature is found to decline in use over time. Diagrammatically this developmental pattern may be represented as follows:

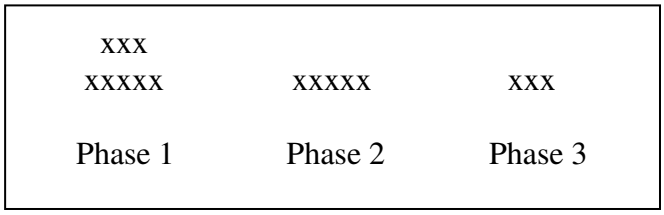


Figure 6.3 A reduction pattern in second language development

In this study, the *many* + category + *of* + n framework (e.g., *many types of flowers*) is a language feature exemplifying this developmental pattern.

4. No consistent pattern

This is a trend with no consistent pattern observable over time. There may be times when a language feature grows while at others it declines in use. In this study, the *v* n *of* n pattern (e.g., *take care of herself*) and the v *of to*-phrase pattern (e.g., *jump of to the river*) are features exemplifying this developmental pattern. It should also be noted that this observed 'no consistent' pattern is based on the longitudinal data used in this study; more data might reveal a pattern that is not observable at the moment.

It is important to point out that all the four developmental patterns can be observed to be working according to the two interactive principles of construction and reduction in second language development (see Chapter 5 for details). Consider the three-phase pattern, for example, with illustrative data from Table 6.18 reproduced below:

From **Table 6.18:** The use of pred-adj *of* n in the ELU data

Time 1	Time 2	Time 3	Time 4
1. full of flowers 2. proud of him (x3) 3. proud of himself (x2) 4. proud of the three boys quick action 5. proud of them (x4) 6. proud of themselves (x2) 7. proud of us	1. full of happiness 2. full of variety types of fishes 3. full of will 4. proud of their act 5. proud of their bravery 6. proud of their son 7. proud of their sons 8. proud of them (x4) 9. proud of themselves 10. proud of you 11. proud of Syaz 12. prouded of Ilman	1. full of fish 2. full of fishes 3. full of worms 4. proud of Daniel's self-sacrifice 5. proud of him (x4) 6. proud of himself (x2) 7. proud of me 8. proud of their son 9. proud of them (x2) 10. pround of himself 11. rich of fish 12. scare of something	1. full of baits and three old rods 2. full of fishes and lilies 3. proud of him 4. proud of their bravery 5. proud of them (x2) 6. proud of themselves 7. proud of you 8. responsible of their river

We find that instances of use of this **pred-adj *of* n** pattern undergo a construction mode of progress (i.e., from seven recurring types of sequences at Time 1 to 12 types of sequences at Time 2 and Time 3), before the reduction principle comes into play. The expansion pattern of development, on the other hand, very much illustrates the construction principle at work while the reduction pattern exemplifies how the reduction principle shapes the process, although it should also be noted that when given a longer time span of observation, these two developmental patterns might exhibit some variation as a result of the two principles working together to model the patterns of development at different phases in time: there may be

features that eventually show 'no consistent' pattern in the developing language. As suggested in the previous chapter, language development entails a constant two-directional process of construction and reduction so that at any point in time, the resources the developing language user enjoys are in a state of flux.

The four developmental patterns observed in this study based on empirical data spanning a period of 24 months can, however, be expected to account for some major patterns observed in language development within a period of study; at the very least, they provide a metalanguage for a consideration of fundamental patterns in the development of a second language.

Overall, the developmental patterns observed in this study indicate that second language development constitutes a highly dynamic process. It is, as shown in the findings of different categories of patterns of use of *of* over time, also an organic process whereby different language features exhibit different developmental patterns. Developing a second language, then, cannot be expected to be a simple process of one's reaching or achieving greater complexity over time; it is a complex process with multiple sub-processes at play involving multiple language features with diverse growth patterns.

The findings on the characteristics and trends of growth of the various uses of *of* examined in this study are summarized and presented in Table 6.20 below.

Table 6.20: Uses of *of* observed in the longitudinal data

	Use at Time 1	Conventional pattern and/or use	Trend of growth	Use at Time 4
1	**n *of* n**	Yes	Increasing	Yes
2	**v *of* n**	Yes	Relatively constant	Yes
3	**v phrasal-prep n**	No	Relatively constant	Yes
4	**v n *of* n**	Yes	Relatively constant	Yes
5	**v n phrasal-prep n**	No	Relatively constant	Yes
6	**v *of that*-clause**	No	Decreasing	No
7	**v *of to*-phrase**	No	Decreasing	No
8	**pred-adj *of* n**	Yes	Decreasing	Yes
9	**attrib-adj *of* n**	No	Decreasing	No
10	**phrasal-prep n**	Yes	Relatively constant	Yes
11	-	No	Increasing	**phrasal conj-subord clause**
12	**Blended pattern**	No	Relatively constant	Yes
13	-	Yes	Increasing	**fixed phrase**

The summary of the results presented here reinforces the notion of the complex nature of the second language developmental process. As can be seen, a total of 13 patterns are observed in the developing language[3], with 11 found occurring at Time 1 and two at Time 4. Almost half of them remain relatively constant in proportion of use, some show an increasing trend of growth while others a decreasing trend of growth. Three patterns (i.e., **v *of that*-clause**, **v *of to*-phrase**, **attrib-adj *of* n**), all of which are innovative, completely die out by Time 4; two patterns absent at Time 1 (i.e., **phrasal conj-subord clause** and the fixed phrase *of course*), one innovative and the other conventional, are, on the other hand, found to be occurring at Time 4.

This study reveals part of ongoing language development, where changes in language use will sometimes represent a move towards convention and sometimes a move away from it. The

[3] A 14th use of *of*, ***of* n**, as shown in Table 6.3 is observed only at Time 2.

development is not entirely linear. Just as the developing language does not become consistently more complex, or less complex, it also does not become consistently more or less conventional.

6.6 Conclusion

This chapter has considered the patterns of use of a function word, *of*, in the ELU language over time. A total of 14 patterns have been identified. Three major categories of patterns, the **n *of* n** pattern, the verb patterns and the adjective patterns, have been studied in detail. The findings suggest that four developmental patterns of language features are observable in second language development, reflecting the trends of how these features grow over time. Such insights are not only compatible with a complex, dynamic, emergent view of language development (e.g., Larsen-Freeman, 2006, 2009; N. Ellis, 2011); they also have the potential to illuminate and explain what a complex process might mean in the context of language development.

This study has demonstrated that second language development is an organic process where changes in usage do not follow single developmental patterns but exhibit substantial variation, and that this variation is not necessarily predictable from looking at the frequency of a single word (such as *of*) only. In fact, a word which appeared unpromising at first, in that its frequency does not change markedly, has led to the discovery of a complex and interesting series of developmental patterns.

This chapter concludes Part II of this thesis, which comprises three studies exploring the process of second language development through function words. The strengths and limitations of corpus studies based on single words will be considered in Chapter 9. In Part III, which comprises two studies, I approach the narratives as written by the ELUs in the longitudinal corpus as individual texts. I consider whether and to what extent individual language users produce longer texts over time, and if there is such evidence, what changes in the structural presentation of narratives might take place through the study of selected individual texts. Together, as noted in Chapter 1, these five studies using tools and methods from corpus linguistics and written discourse analysis represent empirical attempts towards making sense of how second language development takes place over time.

Chapter 7

Analysis of Text Length

7.1 Introduction

One important way to make sense of what second language development over time entails is to consider productivity in language use. To some extent, this was revealed in the previous three studies on the use of *that*, *to* and *of*: there was an increase in clausal control as well as an expansion of lexical resources. In this chapter, I consider productivity purely on a word count basis: how productive have the 124 developing language users been, in terms of increase or decrease in text length, when composing their narratives over time?

This is the first of two studies which bring together investigations into (1) changes in the length of the ELU written narratives over time and (2) how these changes might relate to the ways the ELUs construct their narratives. This chapter focuses on changes in length by examining the distribution of the ELU texts across two time points in 2007 and 2009, while the next chapter considers the relation between increase and decrease in text length and the construction of individual narratives over time.

I begin with an analysis of the distribution of the ELU narratives in 2007 and in 2009. Changes in text length in the narratives are then studied. This is considered from two perspectives: group performance and individual performance. A key finding of this study, as will be shown, is that individual differences in developmental patterns, instead of becoming smaller, turn out to be more pronounced over time.

7.2 ELUs' composition of narratives in 2007 and 2009

As can be seen from Table 7.1, most ELU texts produced in 2007 fall within the relatively narrower range of 90-250 words. This accounts for 90% (or 112) of the 124 narratives produced by the ELUs. A wider range of text lengths were observed in 2009. The same group of ELUs at this later point in time produced a greater number of texts within the range of 151-350 words. 105 narratives (or 85% of the narratives in total) are of this length.

Table 7.1 Distribution of ELU narratives

Text length (words)	Year 2007	Year 2009
90 – 150	36	2
151 – 200	50	34
201 – 250	26	28
251 – 300	8	25
301 – 350	3	18
351 – 400	0	7
401 – 450	0	4
451 – 500	1	6
Total	**124**	**124**

This trend is more clearly visible when presented diagrammatically, as in Figure 7.1 below.

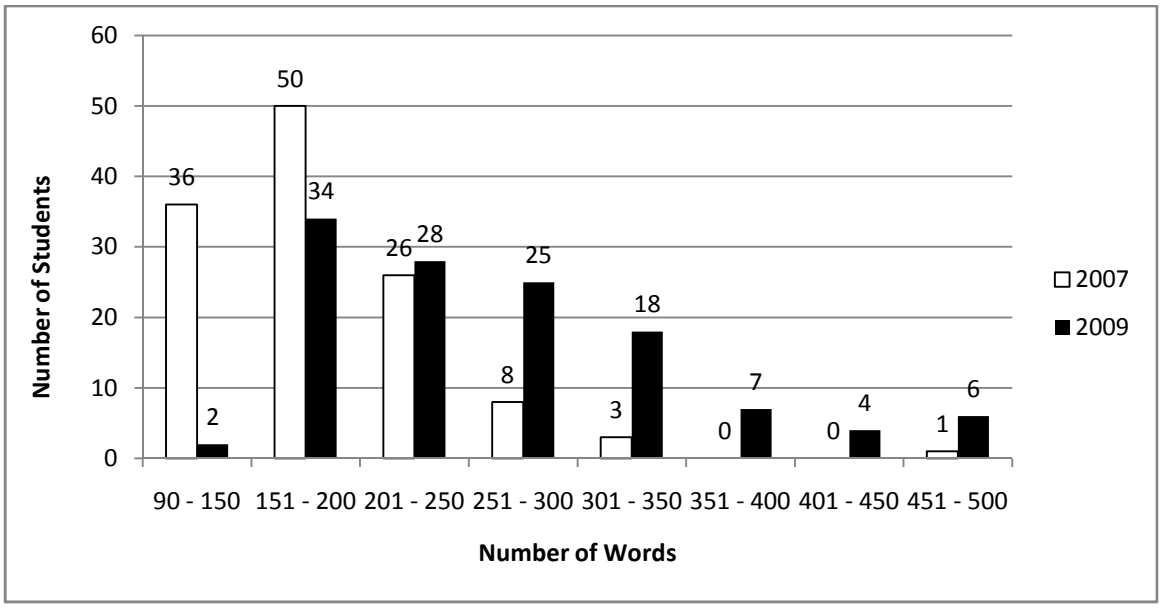


Figure 7.1 Distribution of ELU narratives according to length

In examining productive language use in terms of changes in text length, I was not so much interested in mere increase in the number of words produced by each developing language user; rather, I was more concerned with the *margin of increase* in text length among the ELUs. The difference between the two, in simple terms, may be illustrated by a comparison between an extra 50 words produced at a later point in time by ELU A who had initially

written a 350-word text on the same task, and an extra 50 words produced by ELU B who had written a 100-word essay in the first instance. In this case, ELU A showed an increase in language productivity by 14% whereas ELU B by 50%. That is, it is the amount of change in development that is the focus here.

Table 7.2 presents the results of the analysis of percentage of increase or decrease in text length as observed in the comparison between the narratives produced in 2007 and in 2009 by the 124 ELUs. One striking feature from this table is that most of the developing language users (about 88% or 109 of them) produced longer texts in 2009 than in 2007. Some (19% or 24) of the ELUs produced texts twice as long as those written in 2007, with five ELUs (094, 111, 119, 123 and 090) recording a 200% increase in text length. Equally interesting, a small number of the ELUs (12% or 15 of them) produced shorter texts in 2009.

Table 7.2 Percentage of increase or decrease in ELU text length

ELU ID	2007	2009	%
094	95	347	265.3
111	100	331	231.0
119	100	310	210.0
123	158	478	202.5
090	101	303	200.0
050	162	444	174.1
054	90	224	148.9
127	103	256	148.5
015	138	342	147.8
064	152	366	140.8
051	127	297	133.9
055	147	336	128.6
006	210	476	126.7
010	198	448	126.3
040	102	225	120.6
022	154	332	115.6
020	128	274	114.1
121	180	384	113.3
082	236	500	111.9
036	111	234	110.8
011	163	342	109.8
114	140	291	107.9
095	106	219	106.6

Table 7.2 (continued)

ELU ID	2007	2009	%
086	172	348	102.3
118	129	257	99.2
075	211	420	99.1
101	197	383	94.4
089	254	488	92.1
001	205	393	91.7
067	119	227	90.8
083	251	469	86.9
077	122	226	85.2
026	255	469	83.9
047	102	184	80.4
113	161	290	80.1
002	167	297	77.8
072	200	351	75.5
103	222	384	73.0
070	160	273	70.6
125	178	303	70.2
097	172	292	69.8
106	197	333	69.0
122	116	196	69.0
093	153	257	68.0
059	154	256	66.2
016	124	205	65.3
120	193	316	63.7
053	178	291	63.5
085	202	328	62.4
109	131	211	61.1
017	135	213	57.8
074	185	285	54.1
088	177	271	53.1
065	158	241	52.5
023	127	190	49.6
052	123	184	49.6
080	138	200	44.9
014	134	190	41.8
076	239	338	41.4
100	196	277	41.3
033	171	241	40.9
035	186	258	38.7
045	153	212	38.6
003	127	175	37.8

Table 7.2 (continued)

ELU ID	2007	2009	%
005	142	195	37.3
104	133	182	36.8
024	227	309	36.1
044	115	156	35.7
060	178	241	35.4
056	205	275	34.1
028	174	232	33.3
018	172	229	33.1
038	210	279	32.9
031	134	178	32.8
008	215	284	32.1
062	182	239	31.3
019	139	179	28.8
081	179	230	28.5
034	120	153	27.5
027	147	187	27.2
009	125	158	26.4
087	158	199	25.9
078	208	261	25.5
110	197	245	24.4
091	316	385	21.8
092	202	246	21.8
029	181	220	21.5
098	175	212	21.1
096	265	312	17.7
116	237	277	16.9
048	162	189	16.7
071	171	198	15.8
046	208	235	13.0
049	203	228	12.3
124	206	231	12.1
105	309	345	11.7
057	225	251	11.6
025	182	196	7.7
041	160	172	7.5
084	163	175	7.4
013	218	234	7.3
043	168	179	6.5
102	203	216	6.4
073	249	262	5.2
032	173	182	5.2

Table 7.2 (continued)

ELU ID	2007	2009	%
058	182	190	4.4
079	195	203	4.1
115	150	156	4.0
108	307	318	3.6
030	262	256	-2.3
117	153	149	-2.6
061	190	185	-2.6
069	204	195	-4.4
021	274	256	-6.6
063	171	159	-7.0
126	496	450	-9.3
037	238	215	-9.7
039	212	179	-15.6
112	194	159	-18.0
107	231	189	-18.2
004	226	177	-21.7
066	274	193	-29.6
099	273	192	-29.7
007	200	94	-53.0

The results were further analyzed by comparing growth rates of the top 25 and bottom 25 ELUs from 2007 to 2009, averaged across the data sets (see Figure 7.2).

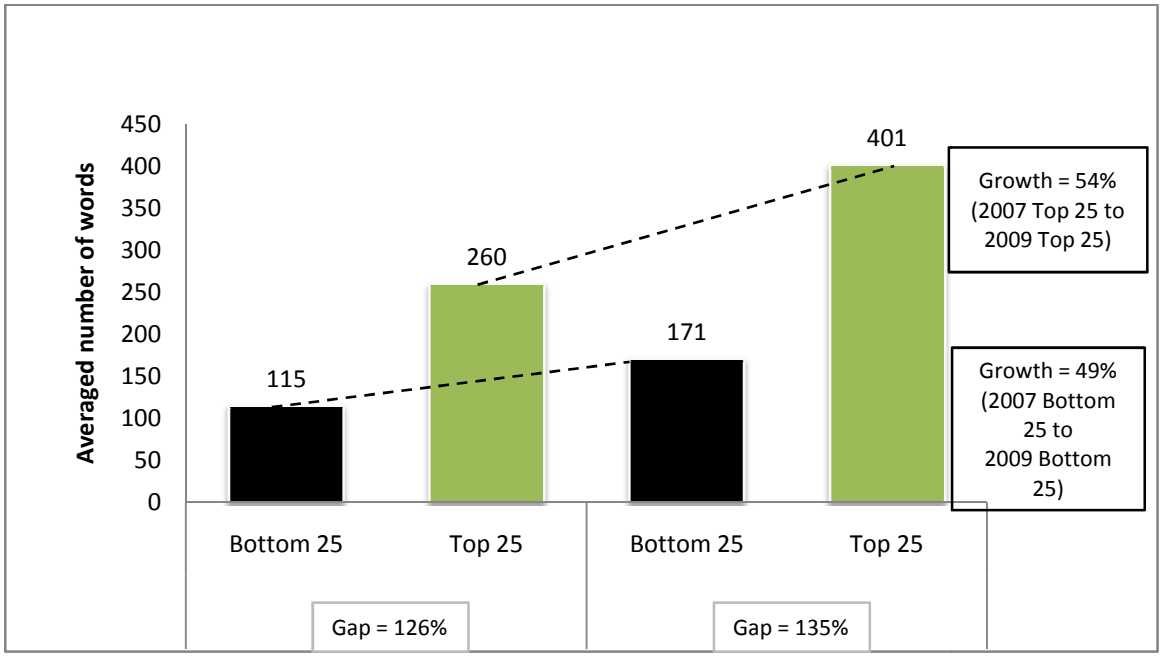


Figure 7.2 Top and bottom 25 groups' production of words averaged across the May 2007 and June 2009 data sets

As can be seen, both the top and bottom groups showed an overall positive trend in average growth. The top groups' average growth, from 2007 to 2009, is 54% while the bottom groups' average growth from 2007 to 2009 is 49%. There is a wider gap between the top and bottom ELU groups as they progressed over the years, with the difference being 9%. While a chi-square test indicates that the change in the gap is not statistically significant ($p > .05$), the rate of change is more pronounced with the Top 25 than with the Bottom 25. The problem is of course there is no information in Figure 7.2 to show whether they were the same developing language users progressing over time: it is not meaningful to talk about 'rate of change' unless we are looking at individual ELUs' development. Any inference made about the rate of change should, therefore, be interpreted with caution.

Essentially this raises the question of how individual ELUs' productive language use actually changes over time, which is considered in the following section.

7.3 Individual ELU performance

As illustrated in Figure 7.3 (see also Table 7.2 earlier), individual performance is more diverse than group performance. While the overall pattern of development suggests a sharp increase in text length over time, there is a wide scattering of individual data, with a small number of ELUs (15 of them), as noted earlier, producing shorter texts in 2009 than in 2007.

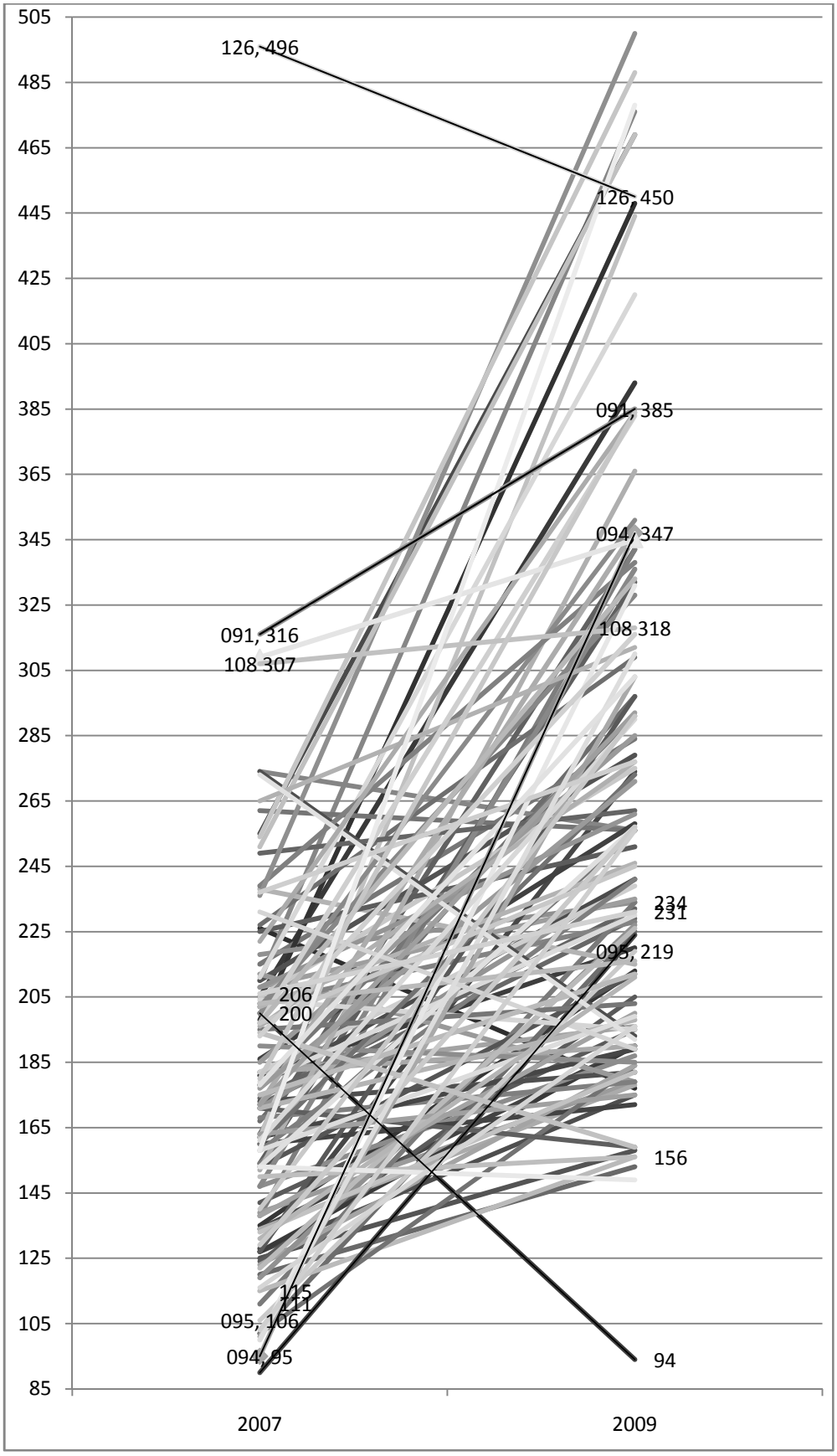


Figure 7.3 Overall trend of text lengths produced by individual ELUs for the years 2007 and 2009

In Figure 7.3, each line on the graph represents one individual ELU. As there is too much information for it to be easily interpreted, Figures 7.4 and 7.5 are used to show the distribution of the number of ELUs according to percentage of increase in text length.

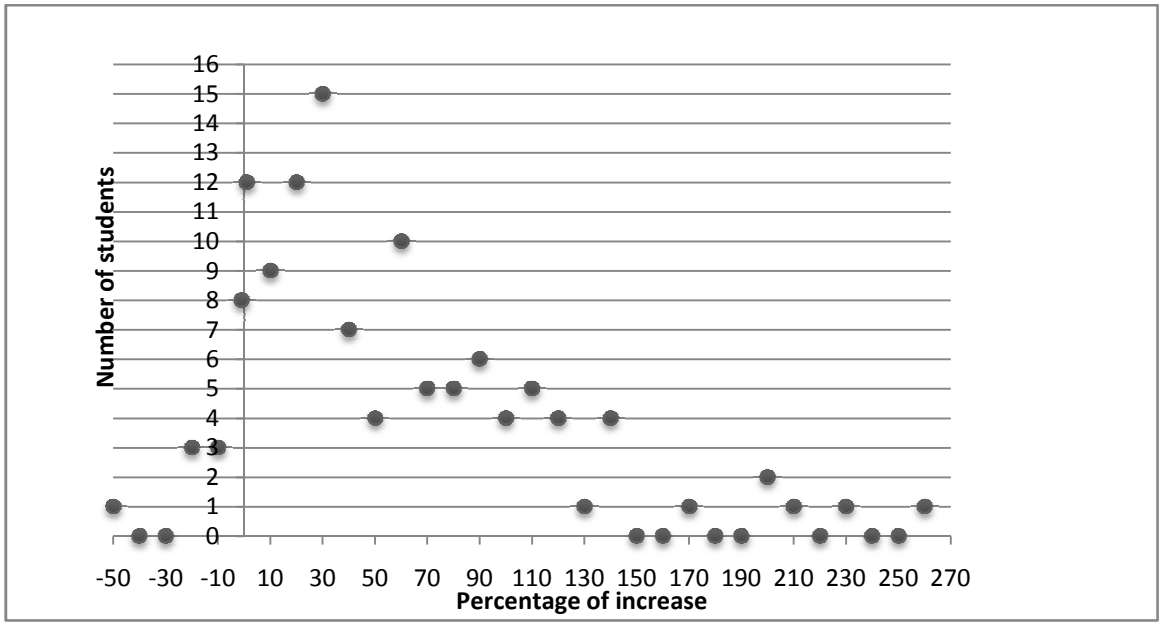


Figure 7.4 Distribution of the number of ELUs according to percentage of increase in text length

Figure 7.4 shows the distribution of the number of ELUs according to percentage of increase in text length in 2009. 12% or 15 ELUs produced shorter texts in 2009 than in 2007. The remaining 88% were observed to produce longer texts over time. Among these, 69% or 85 ELUs produced texts with an increase in length from 1% to 99%. They form the largest group of the developing language users who have written longer texts as time goes by. This is followed by those who increased their text lengths by between 100%-269%. There are 19% or 24 ELUs in this group. All this is summarized in Figure 7.5 below.

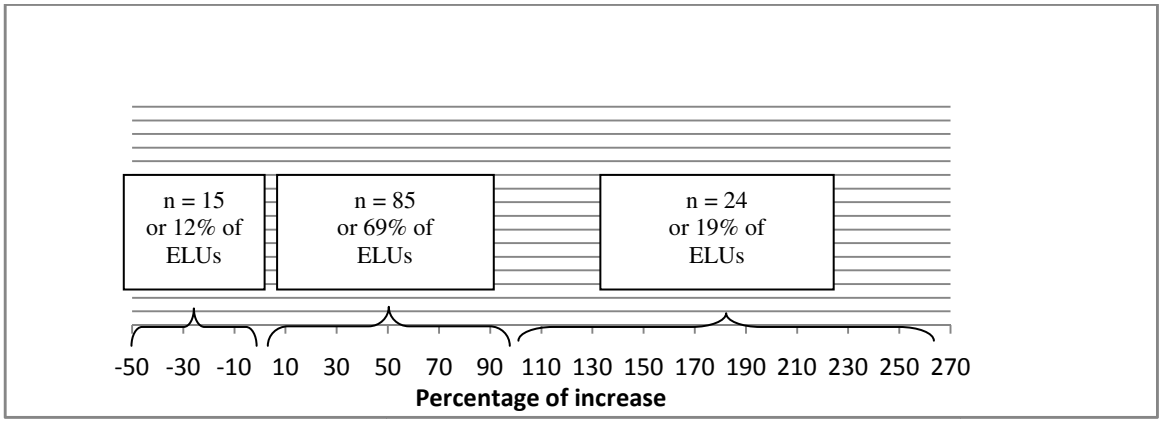


Figure 7.5 Summary of distribution of the number of ELUs according to percentage of increase in text length

It may also be important to point out that within the largest group (69%) who showed an increase in text length, many of them increased by about 30% and the next big sub-group increased by more or less double of what they wrote in 2007.

The mean for the ELU texts produced in 2007 is 180.19 while for those produced in 2009 is 264.13, with the standard deviation being 56.16 and 85.81, respectively. In other words, the 124 ELUs displayed a greater variation in terms of text length when composing their narratives in 2009 than in 2007. Comparisons using the paired-samples t-test found a statistical difference between the two sets of narratives (mean difference = -83.94, 95% CI = -98.78, -69.11, $p < .0001$).

7.4 Conclusion

In this chapter, I have considered how productive the developing language users have been in composing their narratives over a 24-month period based on text length. We may note that the ELUs produce more language (i.e., write longer) over time. We may also note from the findings of the study, however, that it is in many ways inadequate simply to draw such a conclusion. This is because the idea is based on observations of *group* performance; the picture changes when we look more closely at the data, or when we consider the details. A major finding from the study is that the developing language users start off as more or less a homogenous group and become much more differentiated as time goes by. That is, second language development seems to be a process whereby differences in developmental patterns among individual ELUs become more pronounced over time.

The quantitative analyses in this study yield important observations about second language development. There is sufficient evidence both of change and of increase in variability to warrant a more detailed comparison between the narratives written by the same developing language users at different times. A question which can be asked at this point is whether there are changes in the ways ELUs construct their narratives when there is an increase or decrease in text length. I shall, in the next chapter, attempt to address this specific qualitative question: what is observed in the construction of narrative when there is an increase in text length over time?

Chapter 8

A Matrix Study of Narrative Text

8.1 Introduction

As indicated in the previous chapter, there is sufficient evidence both of change and of increase in variability to warrant a closer look at the ELUs' written texts as narratives, and at how different ELUs, and ELUs at different times, construct their narratives differently. The present chapter explores these issues by examining the narratives through a matrix perspective. Originally derived from the work of Pike (1981) and later revised in Hoey (2001), the matrix analysis has been found to be of particular value for the study of narratives, as will be illustrated in the pages that follow.

Three points need to be made before we look at the analysis of the individual narratives. First, while the longitudinal corpus consists of 496 texts, the nature of the analysis, as will be apparent, does not permit discussion of all the ELU narratives. What is to be reported in this chapter are the findings of an analysis of four texts by two ELUs (094 and 054). These texts were chosen from those showing a pattern of increase in text length over time and were impressionistically identified to illustrate typicality of the rest in terms of narrative structure; further research would determine to what extent they were representative but this is beyond the scope of the present study. For reasons of space, those texts by ELUs who exhibit a pattern of reduction in text length as time passes by are not considered, although a brief discussion is presented in Chapter 9 concerning the construction of one sample narrative from these texts.

Second, there have recently been serious research efforts to explore SLA–SLW (second language writing) intersections, as witnessed, for example, in the special issue of the *Journal of Second Language Writing* in 2012. This is an encouraging sign as I, too, share the view that there is much more to learn from ELUs' writing about second language development than was previously assumed; spoken language has been, and still is, the dominant focus of SLA research. The study to be reported in this chapter attempts to contribute to the small but growing body of research (e.g., Hanaoka & Izumi, 2012; Wigglesworth & Storch, 2012; see

also reviews by Bitchener, 2012; Polio, 2012; Williams, 2012) that examines the role of the study of written language in advancing our understanding of language development.

Finally, as with the previous studies reported in this thesis, I am committed to analyzing the texts while rejecting a 'deficit' view of the ELUs. Rather than being compared with an external norm of reference, earlier and later texts by the same narrators are treated as examples of different, changing choices and strategies on the part of the narrators, with, as will be seen, complexity as a distinctive feature emerging from such an analysis. Any 'better' text is observed in the light of 'better than before' or 'better than the earlier one', rather than viewed as 'closer to an idealized norm of practice'. Further, this notion of 'better than before' could just mean that a more complex way of narration is observed. Depending on contexts of use and/or realization, it may, or may not, be an inherent merit in or of itself.

8.2 A matrix perspective on text

In his work on the use of matrices to analyze texts, Hoey (2001) notes that there are several different ways of representing or telling the same happening. The order in which events are recounted, for example, may be different in different tellings. He proposes a matrix analysis of a telling to break down the different elements of a narrative, showing the chronology of events as indicated by the telling, and points out that a different narrative or narratives (tellings) could be derived from the same matrix.

Diagrammatically the relationships among possible tellings may be presented as follows:

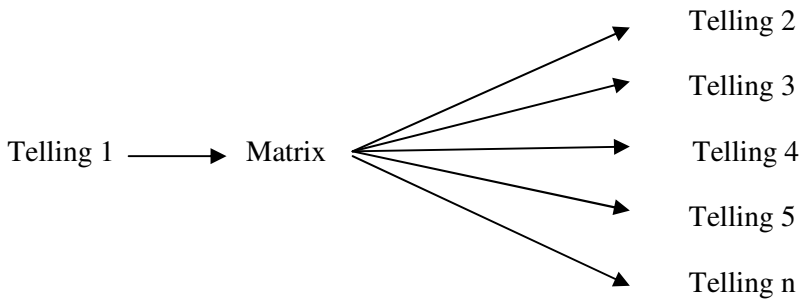


Figure 8.1 From Hoey's (2001, p. 99) revised representation of the relationships among possible tellings

Hoey (2001, p. 99) suggests that a telling always 'precedes and produces our sense of something being a happening' and that alternative tellings are 'presumably mappable onto the

same matrix'. An important point here about the matrix perspective is that different tellings may reveal different paths taken through the same matrix.

This has implications for the analysis of 'learner' texts, as noted by Hoey himself, and for the study of second language development, to be illustrated in this chapter. In the context of the ELU data we have, we may compare the narrative text written by a particular ELU in 2007 with the text written in 2009 by the same developing language user, first by creating a matrix from each telling and then by considering if different paths had been observed in the two tellings and what this may mean in our attempt to make sense of writing development.

8.3 Two ELUs' narratives showing increased pattern of text length over time

This section considers the texts written by two ELUs, 094 and 054, who have composed longer narratives over time. The texts by ELU 094 are examined first, followed by those of ELU 054.

8.3.1 ELU 094: First narrative

Text 8.1 below was written by ELU 094 in 2007; ELU 094 was among the narrators that showed an increase in text length in 2009. In fact her 2009 text is 265% longer than the text she wrote in 2007, as indicated in Table 7.2 in the previous chapter.

Text 8.1 First text by ELU 094 (written in 2007)

Last week, Ahmad, Abu and Ali went for fishing at the river. They brought the equipment of fishing with them. On the way to the river. They saw two girls were plucking the flower near the river. They were Alia and Aiza.

Suddenly, Aiza felt down into the river. Alia excited and shouted Help!Help!. They heard that shouted. They ran to the river.

Ali ran very fast so he jumped into the river and saved Aiza. He took Aiza to the land. Lastly Aiza wake up and said thank you to Ali for helping her.

As can be seen, there are five characters mentioned in the narrative: the boys Ahmad, Abu and Ali, and the girls Alia and Aiza. This, incidentally, corresponds to the number of the characters shown in the prompt (see **Appendix**). We can create a matrix based on the characters noted in the narrative as follows:

Table 8.1 A matrix analysis of Text 8.1

	A **Ahmad**	**B** **Abu**	**C** **Ali**	**D** **Alia**	**E** **Aiza**
1	He went fishing	He went fishing	He went fishing		
2	He saw two girls near a river	He saw two girls near a river	He saw two girls near a river	She was plucking the flower near a river	She was plucking the flower near a river
3					She fell into the river
4				She shouted for help	
5	He heard the shout for help and ran towards the river	He heard the shout for help and ran towards the river	He heard the shout for help and ran towards the river		
6			He jumped into the river to save Aiza		
7			He took Aiza to the land		
8					She 'woke up' and said thank you to Ali

Following Hoey, the horizontal parameter represents the five characters named in the narrative and the vertical parameter represents a series of time frame. The 'time spans', as Hoey (2001, p. 100) points out, can be seen to answer the question of 'What happened next?'. What we have here is a matrix with 40 cells, with labels for the cells derived from the characters and time bands (see Figure 8.2).

	Ahmad	Abu	Ali	Alia	Aiza
Time band 1	A1	B1	C1	D1	E1
Time band 2	A2	B2	C2	D2	E2
Time band 3	A3	B3	C3	D3	E3
Time band 4	A4	B4	C4	D4	E4
Time band 5	A5	B5	C5	D5	E5
Time band 6	A6	B6	C6	D6	E6
Time band 7	A7	B7	C7	D7	E7
Time band 8	A8	B8	C8	D8	E8

Figure 8.2 An abstract representation of Table 8.1

It is not difficult to notice from Table 8.1 that there are some empty cells. The cells A3-A4 and A6 – A8, for example, are empty. So are E4 – E7. What this suggests is the focus of ELU

094 when completing this piece of communicative task in 2007. One could argue that the ELU, if viewed negatively, has neglected to tell us, the readers, what the character Ahmad was doing in time bands A5 – A7, for instance, or what Aiza was doing in time bands E4 – E6. We could, however, also talk about the ELU's point of view at this stage or, perhaps more appropriately, ways of communicating at this point of language development; after all we all display different ways of communicating at different points in time or over time. It is of course also the case that no one, unless deliberately, could tell a story that filled in all the cells: essentially no telling tells everything.

Central to a matrix analysis is a consideration of the path taken through the matrix. The path the ELU's telling takes is shown in Figure 8.3 below. The diagram, which may take a while to become accustomed to, is to be read like normal texts, first horizontally and then downwards. To guide the reading of the diagram, however, arrows have been added to show the 'flow' of the telling. Again following Hoey (2001), the symbol '/' is used to indicate that it is sometimes impossible to separate the telling of two or more cells: certain events or actions take place at or around the same time.

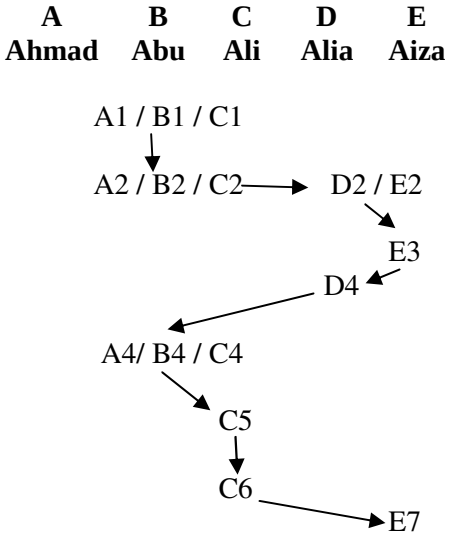


Figure 8.3 The path through the matrix taken by Text 8.1

As far as the path taken is concerned through the matrix, the diagram shows that C is the main character. A and B exist only in relation to him. D exists independently, but only occasionally. E, on the other hand, occurs in interaction with C and D.

The diagram also suggests a somewhat linear progression of narration (note though the arrows are somewhat less helpful to illustrate this; they are there merely to show the 'flow' of the narration). The ELU began the narrative by first focusing on Ahmad, Abu and Ali (A1/B1/C1 and A2/B2/C2) and then on Alia and Aiza (D2/E2). When Aiza fell into the river (E3), her friend Alia shouted for help (D4). The three boys heard her and ran towards the river (A5/B5/C5), before Ali jumped into the river (C6) and took Aiza to the river bank (C7). Finally Aiza thanked him for saving her life (E8). For readers looking for some puzzle-solving pleasure or certain 'suspense' in the story, this piece may be disappointing. To be fair to the ELU, however, she can be seen to be communicatively successful in relaying a story based on the prompt given.

To the SLA or second language writing researcher, what would be more interesting to consider is how the same emergent user might perform when she was responding to the same narrative task at a later point in time. This is what is attempted and discussed below.

8.3.2 ELU 094: Second narrative

Text 8.2 shows the second text written by ELU 094 in 2009. Immediately noticeable is its length; as noted earlier, this text is 265% longer than the first text written in 2007. Would the length indicate or contribute to a different telling of the same story? A matrix analysis of this piece of narrative is presented below.

Text 8.2 Second text by ELU 094 (written in 2009)

Last peaceful Sunday, Ahmad, Naquib and Syakir decided to fishing. They went fishing at wonderful river behind their rustic village. They also known as 'Three Boys'. This is because, wherever they went, they would go together. They also be the apples in the eyes of their teachers. They probably got a great and excellent results in examination.

While they on the way to the river for fishing, they must through along a tiny pond. The pond was usually fulled with the flowers. The place also be famuos to the girls like a beetle see his friends, Alia and Aina. Both of them were also bestfriend, same like Ahmad, Naquib and Syakir. After a few minutes, they continued their journey to the river.

At the pond, Alia and Aina was happily plucking the flowers near the pond. They looked very concentrated to the flowers. While Alia was plucking the flowers, she slippered and felt into the pond. Aina had a big surprised to see Alia in the water. Alia was a chicken hearted, so she can't do anything. She just screamed and make sure Alia still in the water.

Ahmad, Naquib and Syakir heard Alia's screamed and shocked. They worried if Alia and Aina in danger. They were between an evil and deep sea. They didn't know what they should going to. Continued their journey or returned back to Aina and Alia's place? After discussed with each other, they decided to return back to Alia and Aina's place. They had a great surprised when saw Alia in the water like see the big crashing happened in front their eyes.

Naquib who was the great swimmer among the others, ran and jumped as fast as lightning into the water. He tried to get Alia's body and brought her to the land. Alia's condition was not stable yet, she fainted. Aina, who was joined PBSM club, helped Alia with CPR treatment. Thanked to God, it gave a great results. Alia woke up after a few minutes then. And then, Alia thanked to all her friends for helping her.

As can be seen from Table 8.2, there are empty cells in the matrix, which are proportionally fewer in number compared to those in the first text matrix considered above. This is by no means an indication of the second text having ‘more things to say’, however. It is often the amount of detail relating to each cell, rather than the number of filled-up cells, which indicates how much is said. Some cells, as Hoey (2001) points out, contain actions which have to be inferred, and matrices can vary in terms of the level of details the analyst chooses to construct to reflect. Nevertheless we learn something from this matrix analysis: about the redundancy the second text contains and about its absence in the first text, which in turn offers some exciting ways of looking at second language development.

Table 8.2 A matrix analysis of Text 8.2

	A **Ahmad**	**B** **Naquib**	**C** **Syakir**	**D** **Alia**	**E** **Aina**
1	He went fishing	He went fishing	He went fishing		
2	He passed by a pond on his way to the river, and saw Alia and Aina (implied)	He passed by a pond on his way to the river, and saw Alia and Aina (implied)	He passed by a pond on his way to the river, and saw Alia and Aina (implied)	She was (happily) plucking flowers near a pond.	She was (happily) plucking flowers near a pond.
3	He continued his way to the river	He continued his way to the river	He continued his way to the river	She slipped and fell into the pond	
	*In the narrator's words: … *they continued their journey to the river*				
4				She was in the pond	She screamed
5	He heard the scream	He heard the scream	He heard the scream		
6	He discussed with his two friends and decided to return to the pond	He discussed with his two friends and decided to return to the pond	He discussed with his two friends and decided to return to the pond		
7		He ran and jumped into the pond to save Alia			
8		He took Alia to the land		She 'fainted'	She helped Alia with CPR
9				She 'woke up' and thanked all her friends	

The redundancy is particularly pronounced when we consider the path the second text takes through its own matrix (see Figure 8.4):

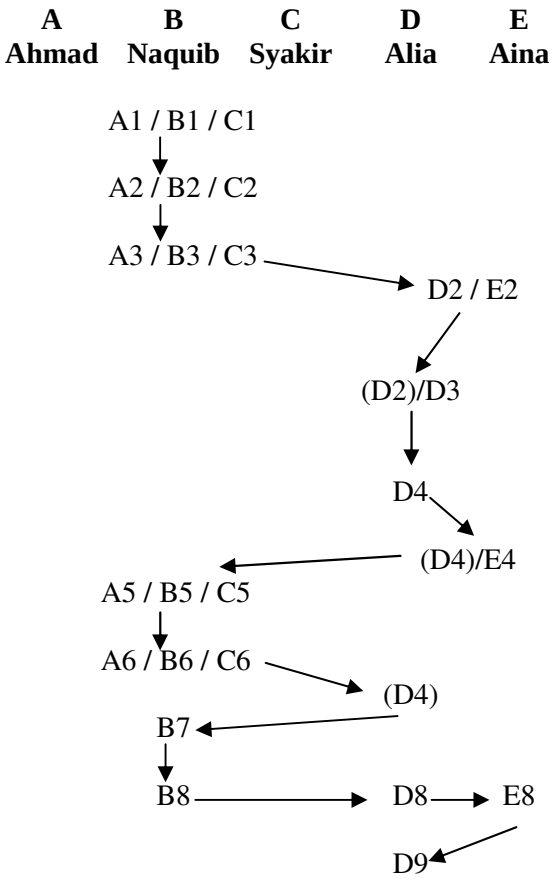


Figure 8.4 The path through the matrix taken by Text 8.2

It can be observed that some cells appear more than once in the path taken by this second narrative, which means we are told the same thing twice or more. In the D column, for example, D2 appears twice and D4 three times, with the symbol '( )' indicating that the idea expressed therein takes its realization in a dependent or subordinate clause. D2 is about the two female characters plucking flowers while D4 about one of them (drowning) in the water or pond. This 'selective' redundancy can be seen to reflect the narrator's emphasis, or even blocks of central 'units of ideas' she used, to give the narrative an overall coherence sense: the prompt is about (perhaps an unfortunate stereotyping of) a drowning girl waiting to be saved.

The cells A6/B6/C6 represent another case of redundancy. Not included in the cells, for obvious reasons of space, is the narrator's description of the three male characters' reaction after they heard the shout for help. In the narrator's words,

> [The three boys] worried if Alia and Aina in danger. … They didn't know what they should going to. Continued their journey or returned back to Aina and Alia's place? After discussed with each other, they decided to return back to Alia and Aina's place. (Text 5.2)

Notice that 'continued their journey' is the same expression used in A3/B3/C3, which is observable two paragraphs away in the written narrative. There are also two instances of 'to return back to Ania and Alia's/Alia and Aina's place' indicated within the same paragraph. While all this may reinforce the idea of coherence in the narrative, especially in the use of *continued their journey* across paragraphs, the notion of complexity also seems to emerge here, a point we will take up shortly.

How do the first and second texts compare? From the matrix analysis, it is clear that the first text displayed a simple, linear narrative progression. In this text, we were told who the male characters were, what they were doing, when they met the two female characters and how, in a rather 'direct' manner, one of the male characters saved the drowning female character, as illustrated in the quoted lines below:

> … They heard that shouted. They ran to the river.
> Ali ran very fast so he jumped into the river and saved Aiza. He took Aiza to the land.
> Lastly Aiza wake up and said thank you to Ali for helping her. (Text 8.1)

The overall linear progression or path taken by the first text is perhaps best captured in the strict sequence of all the cell labels in the matrix as follows:

A1/B1/C1 → A2/B2/C2 → D2/E2 → E3 → D4 → A5/B5/C5 → C6 → C7 → E8

The matrix analysis of the second text written by the same ELU two years later reveals how redundancy was carefully applied to achieve a sense of coherence for the ELU's narration. Central ideas of two female characters plucking flowers and later the drowning female character struggling in the water were noted more than once; so was the expression of the idea of the male characters' contemplating whether 'to help or not to help' when they heard the shout for help.

Particularly important to note is the ELU's latter depiction of the male characters' thoughts through their 'contemplation' act. In the first text, as we have seen, there was only a description of actions by the male characters after they heard the shout for help (*They heard*

that shouted. They ran to the river.). We were not told what the characters thought about or how they felt at that critical moment of the story.

In the second text, however, the male characters were depicted as individuals with their personally developed or developing value systems. The inclusion and depiction of their emotion or mental processes of 'worry' – if the female characters were in danger (*They worried if Alia and Aina in danger*) – and of the need for the characters to decide whether or not to return and help the female character (*They didn't know what they should going to. Continued their journey or returned back to Aina and Alia's place?*), suggests a level of complexity in narration that was absent in the first telling.

The sequence that shows the path taken by the second text of the ELU, which also illustrates the developing complexity of narration at this latter point of language development, is as follows:

A1/B1/C1 → A2/B2/C2 → A3/B3/C3 → D2/E2 → (D2)/D3 → D4 → (D4)/E4 →
A5/B5/C5 → A6/B6/C6 → (D4) → B7 → B8 → D8 → E8 → D9

It should be obvious how different the path taken by the second text is from that taken by the first if we look closely at the part of the sequence after A2/B2/C2 as shown above:

… A3/B3/C3 → D2/E2 → (D2)/D3 → D4 → (D4)/E4 → A5/B5/C5 → A6/B6/C6 → (D4)
→ B7 …

That is, instead of moving linearly on to a time band 4 event after A3/B3/C3, the ELU chose to make a retrospective reference to D2. Also, after describing the event in A6/B6/C6, the ELU returned to the idea expressed in D4 instead of proceeding to a time band 7 event description.

A further note can be made about the sequence towards the end of the matrix analysis which reveals the path taken by, and a distinctive feature observed only in, the second text:

… B8 → D8 → E8 …

This sequence represents the following quoted lines from the second text:

> Naquib … ran and jumped as fast as lightning into the water. He tried to get Alia's body and brought her to the land. Alia's condition was not stable yet, she fainted. Aina, who was joined PBSM club, helped Alia with CPR treatment. (Text 8.2)

What is shown here is simultaneous attention by the ELU in her narration to three different characters, Naquib, Alia and Aina, in the resolution part of the narrative. Both Naquib and Aina were noted to have helped to save unconscious Alia, with Naquib being described to get Alia out of the pond and Aina to perform cardiopulmonary resuscitation.

Compare this with the resolution in the first text, which is noticeably less elaborate and sophisticated, as follows:

> Ali ran very fast so he jumped into the river and saved Aiza. He took Aiza to the land. Lastly Aiza wake up and said thank you to Ali for helping her. (Text 8.1)

To use Hoey's (2001, p. 102) analogy, the first text indicates the work of a narrator who, more or less like the one-handed puppeteer, puts down one character before picking up the other. Over time, however, the narrator was observed to have made an advance in terms of the level of sophistication in control over the change of perspective in narrative. The greater control evident in the second text can be said to represent a notable change for the ELU in approaching the narrative task.

All this shows a choice of narrative focus, control and progression that is overall more complex than a linear order approach the ELU exhibited in her first essay. Implications of findings such as this, in relation to such notions as 'valued' texts (e.g., Kress, 1994), will be briefly considered in Chapter 9.

8.3.3 ELU 054: First narrative

We now turn to the next two texts written by the second ELU, 054, in this study. The first text is shown below, with the matrix analysis presented in Table 8.3 and the path taken by the narrative through the matrix illustrated in Figure 8.5.

Text 8.3 First text by ELU 054 (written in 2007)

One day, I with my friends went to the river for fishing. When we at the river, we saw two girls were picking many types of flower such as lilies and ixora.

Suddenly, we hear a voice, who want to help. Help! Help! We went to the voice and we get one of the girls was drowning. And, me jumping to the river to safe the girl.

She thanks to me because help and safe her from drowning and die. I am so proud because I get safety one people from die.

Table 8.3 A matrix analysis of Text 8.3

	A **I**	**B** **My friends**	**C** **Girl 1**	**D** **Girl 2**
1	I went fishing	They went fishing		
2	When I reached the river, I saw two girls picking flowers	When they reached the river, they saw two girls picking flowers	She was picking flowers	She was picking flowers
3			She fell into the river (implied)	
4			She or Girl 2 shouted for help (implied)	She or Girl 1 shouted for help (implied)
5	I heard the shout for help	They heard the shout for help		
6	I found one girl drowning	They found one girl drowning	She was drowning	
7	I jumped into the river to save the girl			
8			She thanked me for saving her	

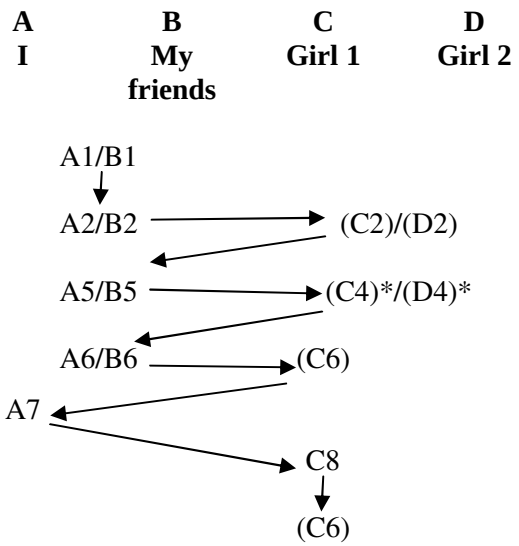


Figure 8.5 The path through the matrix taken by Text 8.3

As can be seen, the ELU begins the narrative by first focusing on 'I' and 'my friends' (A1/B1 and A2/B2) and then on the two girls (C2/D2). There is no explicit mention of which girl had later fallen into the river; what was noted is the male characters hearing the shout for help (A5/B5) by, unspecified by the ELU, one of the girls (hence the use of the symbol '*' in C4*/D4* for this 'unspecificity'). They found one girl drowning (A6/B6 and (C6)) before the character 'I' jumped into the river to save her (A7). She then thanked him for saving her life (C7). While fundamentally linear in approach, what is also notable about this piece of narrative, as observed in a small number of other texts written in 2007 by those ELUs who exhibit an increased pattern in text length over time, is that there is some redundancy found towards the end of the narrative: the idea of drowning (C6) is expressed twice while the act or good deed of saving someone is expressed three times. That is, the whole narrative started off and progressed in a strictly linear manner but rounded off with some redundancy towards the end of the narration. As far as the first texts are concerned, this second ELU, to some extent, exhibited a more complex way of narration than the first ELU we considered earlier.

8.3.4 ELU 054: Second narrative

The second text by ELU 054 written in 2009 suggests a more elaborate treatment of the narrative task (see Text 8.4).

Text 8.4 Second text by ELU 054 (written in 2009)

Last great Saturday evening, Abu, Ali and I went to the longest river in our village. We went there for fishing because there are many fish play in the river. At 5 o'clock in that evening, we walked along the river to find a good place for start fishing. While we are walking, we have see two girls are plucking a bouquet of flower.

After leaving the girls some meters away. Suddenly we heard a thing fell into the river. After that we also heard a girl was shouting loudly for help. After heard the girl shout, we ran as fast as lightning to get the girl and want to know what happen there. Without good thingking, Ali jump into the river after we knew that was happening there and I went to the nearest public phone to call the ambulance and her parents.

After Ali saved the girl and put on the ground, the ambulance that I call just now had been there. The nurse take the girl into the ambulance and sent at the emergency part of hospital. After a few second her parents reach to the hospital.

After the girl get well, her parents were very thanks and gave a reward because had saved their beautiful daugther life. We are very proud for ourselves because had saved a person from die.

A matrix analysis of Text 8.4 is presented in Table 8.4.

Table 8.4 A matrix analysis of Text 8.4

	A **Abu**	**B** **Ali**	**C** **I**	**D** **Two girls**	**E** **The nurse/ ambulance**	**F** **The drowning girl's parents**
1	He went fishing near the longest river in the village	He went fishing near the longest river in the village	I went fishing near the longest river in the village			
2	While he was walking along the river, he saw two girls plucking flowers	While he was walking along the river, he saw two girls plucking flowers	While I was walking along the river, I saw two girls plucking flowers	They were plucking flowers near the river.		
3				One girl fell into the river (implied)		
4	He heard 'a thing' fall into the river	He heard 'a thing' fall into the river	I heard 'a thing' fall into the river			
5				A girl was shouting for help		

Table 8.4 (continued)

	A Abu	B Ali	C I	D Two girls	E The nurse/ ambulance	F The drowning girl's parents
6	He heard a girl shouting for help	He heard a girl shouting for help	I heard a girl shouting for help			
7	He ran to the river to find out what had happened	He ran to the river to find out what had happened	I ran to the river to find out what had happened			
8		He jumped into the river (to save the girl]	I went to the nearest phone booth to ring the ambulance and the girl's parents		They received a phone call to help the drowning girl (implied)	They received a phone about their daughter's condition (implied)
9		He took her to the river bank		She was on the ground	They arrived	
10				She was taken to hospital	They took the girl to hospital	
11						They reached the hospital
12				She became well		They were thankful and rewarded the three boys

One pronounced feature of the matrix shown in Table 8.4 is the expanded number of columns when we compare it with that in Table 8.3. In addition to the 'core' three male and two female characters, the nurse/ambulance and the parents of the drowning girl were introduced in this second narrative.

A recurring feature emerging from a more in-depth analysis, as in the previous analysis of the texts by the first ELU, is the notion of redundancy, again revealed in the path taken by the second text of ELU 054 through its own matrix (see Figure 8.6):

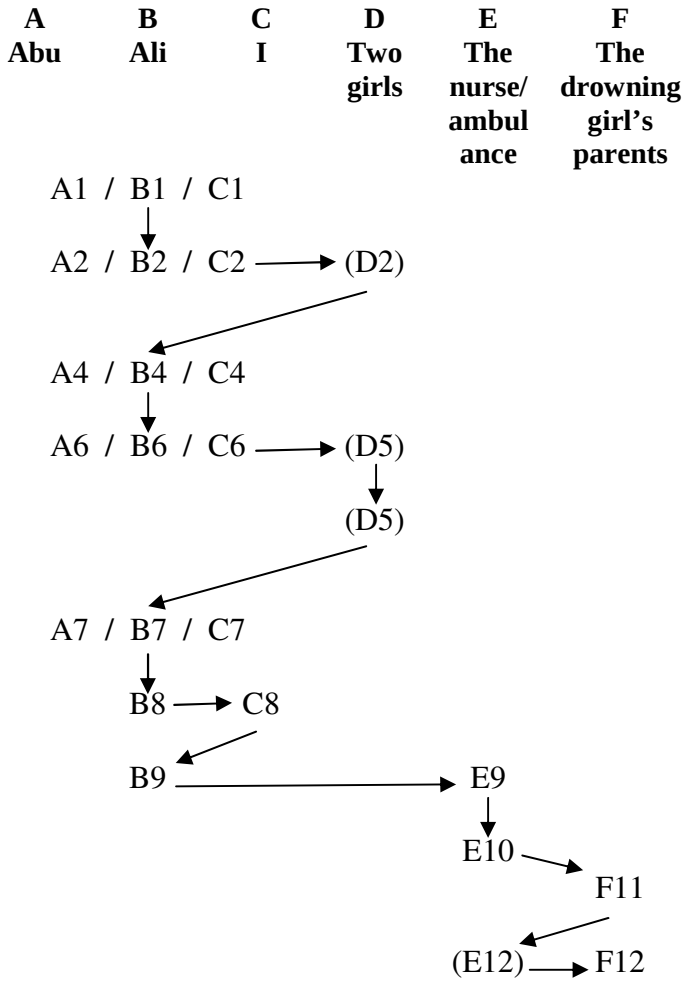


Figure 8.6 The path through the matrix taken by Text 8.4

As can be seen from Figure 8.6, D5 appears twice. It is about a girl's shouting (for help). Also, a reference to the act, or good deed, of saving the girl's life is made three times. Together the use of this redundancy reinforces the central idea of the narrative, which is about a drowning girl and how she was saved.

The cells A7/B7/C7 represent another case of redundancy. Not included in the cells, again for reasons of space, is the ELU's description of the three male characters' attempt to find out what had happened to the girl, or her friend, whose shout they heard. In the narrator's words,

> After heard the girl shout, we ran as fast as lightning to get the girl and want to know what happen there. Without good thingking, Ali jump into the river after we knew that was happening there. (Text 8.4)

When both first and second texts are compared more closely, we find that ELU 054 exhibited some similar strategies of narration in the two texts produced in 2007 and 2009, respectively. Specifically there is a tendency observed in the use of redundancy towards the end of his narrations. A point in case is the expression of the idea of 'saving someone', as noted above, which occurs three times in both texts. Interestingly this expression is embedded only in subordinate clauses in both texts, not found in any independent clause of the texts. Consider the following taken from the first text:

> … And, me jumping to the river to safe the girl.
>
> She thanks to me because help and safe her from drowning and die. I am so proud because I get safety one people from die.

Now consider the following taken from the second text:

> After Ali saved the girl and put on the ground, the ambulance that I call just now had been there. …
>
> After the girl get well, her parents were very thanks and gave a reward because had saved their beautiful daugther life. We are very proud for ourselves because had saved a person from die.

This, incidentally, points to the notion that conducting a matrix analysis is not always a straightforward undertaking. Difficult decisions have to be made in terms of inferring the focus of a sentence and the extent of plausibility of grouping two ideas into one cell or dividing and placing them into two different cells. In a sentence such as *And, me jumping to the river to safe the girl* (Text 8.4), for example, do we devote two cells, one to include 'jumping into the river' and the other to include 'saving the girl'? Or do we group them into one single cell, focusing on the act of 'jumping into the river' with the notion of 'saving' being considered as an extension to the main idea of one's jumping into the river? There is perhaps no right or wrong answer to this. My resolution to this concern as a researcher is, as might be apparent by now, to take note of this feature of redundancy without making the recurring expression of 'saving one's life' explicit either in the matrix or in the path taken through the matrix.

Returning to the matrix analysis of Texts 8.3 and 8.4, overall, we find that both texts make use of similar ways of expressing redundancy, particularly notable towards the end of each

text. The second text, however, additionally complicates the narration through the introduction of a greater number of characters than the first. While the first text is concerned with five main characters as shown in the picture prompt, the second text enriches the narrative with the additional presence of the nurse, ambulance and the parents of the drowning girl.

A further distinctive feature observed only in the second text is illustrated in the following sequence:

… B8 → C8 …

This sequence represents the following quoted lines from the second text:

> … Ali jump into the river after we knew that was happening there and I went to the nearest public phone to call the ambulance and her parents. (Text 8.4)

As in the case with the first ELU whose narratives (Texts 8.1 and 8.2) we have studied, what is shown here is simultaneous attention by the second ELU in his narration to two different characters, Ali and the character 'I', in the resolution part of the narrative. Both Ali and 'I' were noted to have helped to save the girl.

Compare this with the resolution in the first text, which is a one-line description of the saving act, as follows:

> And, me jumping to the river to safe the girl. (Text 8.3)

Hoey's (2001, p. 102) analogy once again applies here. The first text suggests the work of a one-handed-puppeteer-like narrator, who puts down one character before picking up the other. Over time, however, the ELU was observed to be able to display a greater control over the change of perspective in narrative.

8.4 Understanding second language writing development over time

The study reveals five kinds of change that have taken place when text length increases. These are based on the observations of four texts examined in detail and are summarized as follows:

(1) When ELUs produce longer narratives over time, more ideas are introduced;
(2) When ELUs produce longer narratives over time, ideas of a different kind (e.g., thoughts and emotion vs. mere description of actions) are introduced;
(3) When ELUs produce longer narratives over time, a greater amount of redundancy (repeated information) is introduced, with explicit reference ties observed that give a greater sense of coherence and/or emphasis;
(4) When ELUs produce longer narratives over time, a greater number of characters are introduced, with additional information incorporated not only about noting the various characters involved, but also about relating what each did or felt in the narrative; and
(5) When ELUs produce longer narratives over time, the focus of the narrator expands and more than one character is considered at a given time (or 'change of perspective').

Underlying all these changes is the notion of complexity. We have seen, for example, how the second text by ELU 094 represents a considerable advance over her first essay in terms of the manifestation of the characters' thoughts and emotion, in addition to her description of the actions by the characters involved. Complexity in this case can be viewed from two perspectives, as summarized above in (1) and (2). First, when there are more ideas introduced to a piece of writing, the writing becomes more complex. This is a perspective on *quantity* leading to complexity. Second, when ideas of a different nature are introduced (e.g., expressions of the characters' thoughts and emotion versus description of the characters' actions), complexity is recognized here as a matter of quality. This then is a case of *quality* leading to complexity. Both kinds of complexity are observed in the second text by ELU 094, and to a lesser extent, by ELU 054.

Complexity also encompasses the idea of redundancy, as noted in (3). This is particularly true in the case of the first ELU (094). As shown in the matrix analysis, the path taken by the second text of the ELU is different from that taken by the first text. While the first text records events as they are seen from the prompt and describes them in strict sequence, the second text, while still largely chronological in organization, can be seen to make more

explicit reference ties within and across paragraphs throughout the text. Complexity in narrative progression emerges as a result, and the 'redundancy' approach has a demonstrable effect on the text length.

In addition, complexity arises as a result of having 'more characters' introduced in the narrative. When there are more characters, naturally there are more roles to be distributed and related. The second text written by ELU 054 best exemplifies it. By introducing additional characters into the second text (compared to his first text), the developing narrator has evidently expanded his second text. It is important to note, however, that introducing complexity by manipulating a greater number of roles or characters in a narrative may take place at the expense of introducing complexity from a redundancy perspective.

The qualitative difference in narration between the second text by the first ELU (094) and the second text by the second ELU (054) may illustrate this. While ELU 094 produces a longer text over time largely by exploiting such redundancy resources as repeated information (i.e., without recourse to more characters), ELU 054 increases his text length over time mainly through introducing a greater number of characters in his narrative (i.e., without a substantially greater use of redundancy). There could be a trade-off effect, then, between an attempt to exploit redundancy on the one hand and an attempt to enrich a narrative with more characters on the other. Either way, complexity has been observed to emerge from such efforts.

Finally, different levels of control over the change of perspective, which also suggest different levels of complexity, may correspond to changes in text length. The first text by both ELUs runs what Hoey (2001, p. 117) might call the problem of an overconsistent perspective. From the matrix analysis, we observe how the linear sequence of progression in the text often entails the presence of one character or one group of characters at a time throughout the text. Thus when a boy was described to be saving a girl, for instance, the other boys and the girl's friend were not mentioned or assumed to be watching the saving act. The result is a considerably shorter text, standing in contrast to the second, longer text.

As we have seen, both the ELUs in this study have, at one point in their second texts, paid simultaneous attention to what different characters were doing in their respective narrative. This is a level of complexity absent in the first texts by the two narrators. Of course there is

nothing necessarily wrong with consistency in perspective in a narrative: as pointed out by Hoey, many well-known narratives such as *Robinson Crusoe* are written from few perspectives. For the ELU texts we have studied, however, a more sophisticated control over the change of perspective can be observed when the text is longer, leading to the kind of complexity noted in (5) above.

8.5 Conclusion

This chapter has examined four narrative texts composed by two ELUs over time. The findings are revealing of the different types of complexity emerging as a result of an increase in text length over time. Essentially a greater text length means more complexity, which is about additional information or repeated information being introduced into the text. The additional information may, in turn, be about 'more of the same' or 'different kinds'. The introduction of more ideas, the incorporation of a greater number of characters in the narrative, the projection and integration of different ideas, and the changing of perspective all play a role in contributing to the observed changes in complexity of a text. Second language writing development, as suggested in this study, seems to be a process of becoming more complex over time. The extent to which this is generalizable awaits confirmation though in future work that employs texts which are produced across different points in time.

This study marks the end of a total of five studies presented in this thesis. In the next two chapters, I seek to answer the research questions posed in Chapter 1.

Chapter 9

Discussion

9.1 Introduction

This chapter and the next attempt to address four questions initially raised in Chapter 1. The questions are:

1. How does the use of three selected closed-class words among the ELUs change over time and what does this tell us not only about those words but also about other aspects of the developing language?
2. (a) As ELUs repeat the narrative task over two years, do they write more or less?
   (b) Does the structure of the narrative remain the same or does it change? If it changes, in what way does it change?
3. Given the ideological standpoint that I approach the task of researching the developing language with respect, for it to be viewed as another form of natural, human language in its own right and not against a native-speaker norm or an external point of reference, is it possible for me, the researcher, to remain true to that standpoint while I am studying the changes in the developing language that take place over time?
4. What does the study of a longitudinal corpus add to the literature of learner corpus research and to our understanding of second language development, including second language writing development, over time?

In this chapter, I focus on the first two questions, leaving the remaining questions to be addressed in Chapter 10. An account of evaluation is also presented of the approaches to data analysis adopted in this thesis.

9.2 Research Question 1

> How does the use of three selected closed-class words among the ELUs change over time and what does this tell us not only about those words but also about other aspects of the developing language?

Using the function words *that, to* and *of* as the starting point in a series of studies to explore the developing language in LoCDeLUNT, I have adopted an approach that focuses not on any specific part of the lexical or grammatical resource of the developing language, but rather allows for the single words to lead the analyst to uncover the underlying workings of the language. In casting a wide net in searching for signs of change in the developing language, this thesis has noted a number of observations.

These observations are presented below, with the discussion organized as follows:
(1) Change in relative frequency of use of *that*, *to* and *of* over time
(2) Change in patterns of use of *that*, *to* and *of* over time
(3) Changes in other aspects of the developing language over time

9.2.1 Change in relative frequency

In researching the use of the word *that* among the ELUs as reported in Chapter 4, it has been observed that this word occurs as the 35th most frequently used word at Time 1 of the longitudinal data. This relative frequency of use changes at Time 2, where *that* was found to be the 20th most frequently used word in the developing language. This increase in relative frequency of use continues at Time 3 and Time 4, with the function word seen to occupy the 15th place and the 11th place, respectively, in the computer-generated frequency list. An increasing, proportional use has thus been found of the word as time goes by. This finding based on the longitudinal data is reinforced in the study of the cross-sectional data in which *that* only appears in the top 20 high frequency word list of the texts written by the oldest ELU group (i.e., the 16-year-olds).

While *that* shows an increase in proportional use over time, the next function word considered in this thesis (Chapter 5), *to*, shows consistency in its relative frequency of use across the four points in time in the longitudinal study. It is consistently ranked as the second most frequently used word from Time 1 to Time 4 in the developing language. The same finding is observed in the cross-sectional data: *to* is the second high frequency word used by the 11-year-olds, 13-year-olds and 16-year-olds. The findings from the analysis of both the longitudinal data and the cross-sectional data have, in these two studies, pointed to a convergence of evidence in language use.

The final word examined in this thesis (Chapter 6), *of*, shows no consistent change in the relative frequency of use in the developing language. It is the 13th high frequency word observed at Time 1 and 17th at Time 2 before an increase in its relative frequency to 13th was observed at Time 3, and a decrease to 16th at Time 4. This longitudinal picture of developmental pattern is, however, not observable in the cross-sectional data. In the latter data, there is a clear increase in both absolute and relative frequencies in the use of *of* by the younger ELUs to the older ELUs. The word ranks number 19 in the relative frequency of language use by the 11-year-olds, 13 by the 13-year-olds, and nine by the 16-year-olds, thereby raising questions of generalizability and applicability of findings from studies based on a cross-sectional design to inform our understanding of language development over time.

9.2.2 Change in patterns of use

There are four categories of use of the word *that* which have been observed in the longitudinal data:

1. *that* as a determiner
2. *that* as a demonstrative pronoun
3. *that* as a relative pronoun
4. *that* as a conjunction

They have further been grouped into two broader classifications of language use: phrasal *that* constructions and clausal *that* constructions. The former consists of the categories of *that* as a determiner and *that* as a demonstrative pronoun, signifying usage within phrasal units. The latter, on the other hand, consists of the categories of *that* as a relative pronoun and *that* as a conjunction, indicating more complex language use involving the linking of clausal units. The developing language users have been found to rely, initially, more heavily on phrasal *that* constructions in their composition of narrative texts, before shifting to more extensive use of clausal *that* constructions at a later point in time in the construction of their narratives.

Similarly, a total of four categories of use of the word *to* have been observed in the longitudinal data:

1. *to* as an infinitive marker
2. *to* as a preposition
3. *to* as an adverb

4. *to* as an Other clause marker (i.e., followed by a verb ending with *-s/-ed/-en/-ing*, or a word which takes the form of a noun, an adjective or an adverb, to construct a clause)

As in the study of *that*, these four categories of *to* have further been grouped into two broader classifications of language use: intra-clausal constructions and inter-clausal constructions. The former consists of the categories of *to* as a preposition and *to* as an adverb, denoting usage within clausal units. The latter, on the other hand, consists of the categories of *to* as an infinitive marker and *to* as an Other clause marker, suggesting more complex language use involving the conjoining of clausal units. A main finding of this second study is that the inter-clausal constructions have formed a slightly more dominant area of use in the developing language than the intra-clausal constructions since Time 1, and this area of language is observed to continue to expand in proportion of use in the crafting of the narrative texts among the ELUs as time goes by. Another finding of this study is that the infinitive marker steadily increases in proportion across the four points in time over the 24-month period while the Other clause marker is found to be decreasing in proportion over time.

Turning now to the final corpus study of this thesis which is on the use of *of*, we find a total of 14 categories of its use in the developing language:

1. A noun is followed by *of* and a noun phrase: **n *of* n**
2. A verb is followed by *of* and a noun phrase: **v *of* n**
3. A verb is followed by a phrasal preposition ending with *of* and a noun phrase: **v phrasal-prep n**
4. A verb is followed by a noun and *of* and a noun phrase: **v n *of* n**
5. A verb is followed by a noun and a phrasal preposition ending with *of* and a noun phrase: **v n phrasal-prep n**
6. A verb is followed by *of* and a *that*-clause: **v *of that*-clause**
7. A verb is followed by *of* and a prepositional *to*-phrase: **v *of to*-phrase**
8. A predicative adjective is followed by *of* and a noun phrase: **pred-adj *of* n**
9. An attributive adjective is followed by *of* and a noun phrase: **attrib-adj *of* n**
10. A phrasal preposition ending with *of* is followed by a noun phrase: **phrasal-prep n**
11. A phrasal subordinating conjunction ending with *of* is followed by a clause: **phrasal conj-subord clause**
12. *of* is followed by a noun phrase: ***of* n**
13. Blended patterns:

(a) ***a*** **n/adj** ***of*** **n**

(b) ***the*** **n/v** ***of*** **n**

(c) **adv adj/n** ***of*** **n**

(d) **prep n/adj**

(e) ***to*** **n/prep** ***of*** **n**

(f) **pred-adj** ***of*** **n/v**

(g) **pred-adj** ***of*** **n/adj**

14. fixed phrase

What is shown in this study concerning change in patterns of use over time is overall less clear-cut than those observed in the previous two studies. This is particularly true in the general finding of the use of different categories of *of* over time: there are no notable differences as far as change in proportion of use is concerned of the categories identified. What seems to be the least informative study based on the quantitative analysis of a function word, however, turns out to be most revealing about patterns of development in the emergent user language. This is especially true when the qualitative aspects of the developing language are also taken into consideration. As will be made clear in the next chapter when we approach Research Question 4 on the contribution of the study of a longitudinal corpus to our understanding of second language development, this study has demonstrated that changes in usage do not follow a single developmental pattern; they manifest considerable variation. Further, this variation is not readily accounted for by the study of the frequency of a single word alone such as *of* : the frequency does not increase but the variability does. In other words, observations about frequency do not necessarily correspond to observations about usage.

9.2.3 Changes in other aspects of the developing language over time

A number of unexpected observations have arisen from the three studies on function words. These include the observations of an ongoing, dynamic interaction between lexis and grammar in the developing language, of a reduction mode of progress involving recurring linguistic patterns, and of a change in the meaning-making capacities among the developing language users.

9.2.3.1 Dynamic development of regularities of lexis-grammar association

One important finding of the research reported in this thesis is about the changing configurations of the linguistic resources comprising lexis and grammar in the developing language. The study on *that*, for example, has shown that an increase in clausal *that* constructions in the developing language over time comes in company with a wider range of lexis in the longitudinal data.

At Time 1, for instance, the lexis observed occurs in the categories of nouns and verbs. The nouns occur with relative clauses such as the following:

(1) *The flowers* that she hold also fell into the river. (015a)
After arrive at the river, they tried to found *the place* that be suitable for they go to fishing. (077a)
Unfortunately, *the three boys* that want to fishing hear the voice for help. (096a)

The verbs, on the other hand, are observed to be associated lexis co-occurring with *that*-complement clauses, together forming a converging pattern which has been identified in Francis et al. (1996) as the verb + *that* pattern:

(2) They *said* that they saved Manie because she was their friend, not because of the money. (105a)
The doctor *says* that Sarah must have more rest. (103a)
She *promised* that she will be careful later and Balqis thanked to her friends Diana because her helped. (076a)

At Time 4, an expansion of lexis is observed in the developing language. The list of nouns has expanded; so has the list of associated verbs. In addition, lexis in the category of adjectives makes its appearance at this point in time, and this is associated lexis co-occurring with *that*-complement clauses, forging a converging pattern which has been identified as the adjective + *that* pattern in Francis et al. (1998):

(3) Samad also *happy* that he was helped that girl. (036d)
All of them were very *happy* that Lucy didn't face any critical injury.
When they reached, they were *shocked* that one of the girls was fall down into the pond when she was plucking flowers. (045d)

All this suggests how quantitatively and qualitatively the developing language users change their language resources as they grow with greater experience over time. There are attested,

substantial differences in the choice and use of lexis in the developmental linguistic repertoires. The evidence presented in this study points to the important observation that lexical and grammatical properties constitute dynamic co-emergent building blocks of language development.

The final point on the development of regularities of association between lexis and structure is reinforced and confirmed in the next study on the use of *to* among the developing language users. The increase in proportion of use of the inter-clausal *to* over time has been found to take place at the same time an expansion of lexis is observed.

At Time 1 of the developing language, the emergent lexis occurs in four categories: noun groups, verb groups, adjective and adverb groups and an 'Others' category, consisting of words such as *about* and *can*. Similar to the findings of the first study, the nouns are observed to occur with clauses of purpose with *to* such as the following:

(4) They also gave the boys a *present* to show how much they thanked them. (056a)
Dina is very panic and call the three *boys* to help her …. (108a)
Mr Jamal and Nana's brother jumped into the *river* to saved Nana. (110a)

As can be seen from the examples here and from Table 5.3 presented in Chapter 5, none of the nouns and the instances of clauses of purpose occurring at Time 1 form mutually constitutive patterns, as realized in the notion of pattern grammar (Francis et al., 1996, 1998). Associated lexis is, however, observed at this point of development to occur in the categories of verb groups, adjective and adverb groups and the 'Others' category. It co-occurs with clauses of complementation to form the verb + *to* pattern, the adjective + *to* pattern and the fixed phrase of *about to*, as follows:

(5) John quickly jumped into the river and *tried* to saved Elena. (081a)
They felt *shocked* to see a girl was drowned while her friend was shouting for helped. (088a)
Her friend are *about* to drown. (066a)

At Time 4, a substantial expansion of lexis is observed in the developing language. The size of proportional difference is most pronounced in the categories of verb groups and adjective and adverb groups. In addition, one recurring noun phrase (*a good idea*) in the category of noun groups makes its appearance at this point in time, and this is an example of associated

lexis co-occurring with clauses of complementation, forming an innovative pattern specific to the developing language in focus (i.e., absent from the description, but in line with the working principle, of pattern grammar (Hunston & Francis, 1999):

(6) Amri and Chong thought that was *a good idea* to go for fishing during school holidays. (082d)
She thought that *a good idea* to save her beloved friend. (120d)

Further, a greater number of complex patterns comprising three or more elements are observed at Time 4. These include the complex verb patterns exemplified in the verb + noun + *to*-infinitive (7), and the emergent adjective patterns with *it* in the sequence of *it* + be + adjective + *to*-infinitive (8):

(7) He *tells Amer and Merul to confirm the incident*. (013d)
The also *advised me to be more careful* next time when Hana and I wanted to pluck flowers again. (085d)
My friend *asked me to saved the girl*. (99d)

(8) *It is hard to swim and save someones* with shirt. (082d)
Lina had advised her do not pluck the flower because *it was to dangerous to get them*. (118d)

In short, the studies conducted through the exploration of single function words, *that* and *to*, reveal how the developing language is becoming richer and more sophisticated with an ongoing, remarkably dynamic emergence of linguistic resources that comprise lexis and grammar.

9.2.3.2 Declining use of patterns

While a growth in the range of lexical and grammatical resources has been observed in some contexts, it should be remembered that the opposite is also true: some resources fall into (comparative) disuse over time. This is most noticeable in the study based on the word *to* in the developing language. Patterns and sequences such as modal + (negative +) *to*-infinitive (e.g., Mary was afraided but she *did not to do* now), noun + *to*-infinitive with *there* to indicate existence rather than purpose or need (e.g., There are *three boys at the lake to go fishing*) and *know* + *to*-infinitive (e.g., She doesn't *know to swim*), for example, have been observed to be declining in use over time. In fact, two of these three types of sequences (noun + *to*-infinitive

with existential *there* and *know* + *to*-infinitive) are found to have been completely 'died out' by Time 4.

A major category of use of *to* which shows a declining trend of growth in the study is the Other clause marker. As noted in Chapter 5, this is a feature of innovative language use, which can be further classified into two different usages. The first is that *to* is observed to be followed by a verb ending with *–ed, –en*, *–ing, –s* or with an irregular form, to construct a clause, as illustrated in the following examples:

(9) Shakila asked Farisha *to joined* her picked up some flower to bring home. (047c)
The girl who felt into the river was did not know how *to swimming*. (077c)
They loved *to went* there because the lake was so beautiful. (116a)

The second usage of *to* in this Other clause marker category is observed to be followed by a word that takes the traditionally perceived form of an adjective, an adverb or a noun to form a clause, as follows:

(10) Tina was promised *to more carefull* went at the side of river. (121a)
Finally, Fauzana was manage *to still alive*. (058d)
I told to Rahim that he have became a hero today and he just smiled while tired *to breath*. (026d)

The frequency of the Other clause marker accounts for 13.8% of all instances of use of *to* at Time 1 of the developing language. It is observed to be decreasing in use over time, falling to 8.0% at Time 4.

In Chapter 6, the *many* + category + *of* + n framework has been noted as an example feature exhibiting a gradual, declining use. Initial points of the developing language witness a range of alternative expressions such as *many types of flower, many kind of flowers, many type of flowers* and *many types of flowers*. At the later points of development in the study, however, only the expression *many types of flowers* remains in the language. This is but one of a number of other patterns of use observed with a relatively low frequency of occurrence which have been documented in Chapters 5 and 6. Together they show a declining trend of growth of the linguistic features observed in the developing language. It must be noted though at this point that the number of their occurrences together with the changes in their frequency is generally small. Future research might pursue further this promising line of enquiry.

Overall through the studies of three function words, we find not only an expansion of lexical and grammatical resources over time but also a simultaneous reduction in the number and frequency of linguistic patterns of use in the developing language. Each of these findings is important and will be taken up again and elaborated further in the next chapter when Research Question 4 is addressed.

9.2.3.3 The meaning-making dynamics of the developing language

The most unexpected results, however, based on the series of studies reported in this thesis, at least to me as the researcher, relate to the observations about how the developing language constitutes a dynamic meaning-making resource.

That is, when setting out to conduct the analysis of the emergent user language through the three function words, what I had in mind and what I expected to be an outcome of my research was a series of findings concerning the structural aspects of the developing language, which would, in my mind at that time, be welcoming and respectable findings in their own right. As my analysis of the corpus showed a more extensive use of lexical resources with an increased range over time, co-occurring with an observed, greater use of *that* and *to* and together forming regularities of language use, however, my own view began to change. I started to look at the object of my investigation in a new light: the longitudinal corpus, recording four relatively extended moments of language use by the secondary school students over two years in my research, is not merely a collection of visible strings of words with abstract structures.

The systematically changing patterning of lexis and grammar uncovered from the corpus, much of which conforming to aspects of findings obtained from the analysis of a different and larger corpus as documented in Francis et al. (1996, 1998), is a particularly important empirical finding: across time and space, there is converging language use attested in samples of the developing language, collected between 2007 and 2009, by a specific group of students in Malaysia and in samples of language, collected between 1980s and 1990s, by users of English mainly from the UK and the US.

What could possibly be binding such broadly convergent language use? As shown in the studies reported in Chapters 4 and 5 (and to be further discussed below), there is a noticeable increase in the number and range of meaning categories or groups observed in the developing

language as time goes by, with the emergent lexis found to be largely corresponding to that recorded in Francis et al. (1996, 1998). The developing language is, in other words, not just a structural system: it is an ongoing meaning-making resource. One direct outcome from this realization is a revision of an earlier draft of Chapter 1 of this thesis to include, necessarily so, the notion of the developing language as a 'dynamic meaning-making resource' and to refer to the developing language users in this thesis as 'dynamic meaning makers'.

Meaning-making through the lens of the function word *that*

We might recall that in the study on *that*, for example, there are only three different instances of noun groups (i.e., *the flowers, the place*, *the three boys*) observed at Time 1 and these noun groups are reflective of only one type of nouns: the concrete nouns. At Time 4, however, a total of sixteen instances of noun groups have been found to be recurring in the developing language (i.e., *friends, pond, something, story, the boys, the flower, the flowers, the girl, the place, the river, the shouted, the three boys, them, two girls, voice, experience*). As can be seen, both concrete and abstract nouns are observed to be emergent meaningful units at this point of development.

A clearer expansion of instances such as these small, meaningful units of language use which form part of the larger events or units of meaning (Sinclair, 1991, 1996) can be witnessed in the use of verb groups by the developing language users. At Time 1, only three different recurring instances (i.e., *said, says, promised*) of verb groups denoting processes of communication are observed. As time goes by, however, this number has increased to nine (i.e., *promised, said, she said, the doctor says, told, told her, told her parents, told [proper noun], told to*). This is not the only observation at Time 4; instances of two new kinds of verb groups with *that*-clauses are found to emerge at this point in time: those denoting mental processes (i.e., *hope, hoped, knew, knows*) and those denoting discovery processes (i.e., *found, found out, heard, noticed, realize, realized, they saw, to see*). Further, a new group of meaningful units, the adjective group, makes its appearance at Time 4 with *that*-clauses. This group comprises the recurring instances of *happy* and *shocked*.

What we find from the study of *that* then is evidence of how the developing language users are expanding their meaning-making capacities observable in, through and with the change in their descriptions over time of the characters, activities and events in their narratives. At Time 1, with the dominant recurring use of meaningful units surrounding concrete nouns and verbs

of communication with *that*, the narrative world constructed appears to be one where there are only physical objects (i.e., *the flowers, the place*, *the three boys*) in place and where communication (i.e., signaled by such verbs as *said, says, promised*) is the only one type of activity or event taking place.

At Time 4, however, the narrative world has been painted with a very different picture. In this newly constructed world, not only are there more physical objects in place (i.e., *friends, pond, the boys, the flower, the flowers, the girl, the place, the river, the three boys, them, two girls*), a larger, more varied set of communication events (i.e., *promised, said, she said, the doctor says, told, told her, told her parents, told [proper noun], told to*) has also been observed to take place between characters in the narrative. In addition, it should be pointed out that the inner mental processes of the characters (i.e., *hope, hoped, knew, knows*) have been related. Perhaps most important of all, the characters have been portrayed not only as thinking and sensory beings (i.e., *found, found out, heard, noticed, realize, realized, they saw, to see*), but also as beings with feelings (i.e., *happy, shocked*). It is probably not too exaggerated a claim to make, therefore, that a transformed narrative world has been constructed, signifying the developing users' active engagement and continuous development in their resources for meaning-making.

Meaning-making through the lens of the function word *to*

This set of findings gains further support from another study based on the use of *to* in the developing language. This second study has, in fact, shown a far larger production of meaningful units of language use. Interested readers are referred to Tables 5.3 and 5.4 in Chapter 5 for details of the individual units identified. To be discussed here is a smaller list of the units but one which is itself reflective of the overall trend of the relevant findings, with each recurring instance observed to be combining with a *to*-complement clause to form a larger meaningful unit of language use. This list is reproduced below, from Table 5.5. Those items underlined are innovative instances or categories of units as identified and explained in Chapter 5.

From **Table 5.5:** Lexis associated with *to*-complement clauses at Time 1 and Time 4

		Time 1	Time 4
Noun group	1.	–	The 'Thought' group a good idea
Verb groups	1.		*The 'Begin' group* started
	2.		*The 'Discover' group* learn how
	3.	*The 'Help' group* helped [somebody]	*The 'Help' group* helped [somebody]
	4.	*The 'Hope' group* do not want, want, wanted, wants	*The 'Hope' group* want, wanted, wants
	5.	*The 'Like' group* like	*The 'Like' group* like love/loves
	6.		*The 'Manage' group* manage/managed
	7.	*The 'Promise' group* planned	*The 'Promise' group* decide, decided plan, planned, planning promise, promised
	8.	*The 'Remember' group* did not know how/didn't know how do not know how/ don't know how does not know how/doesn't know how know what know didn't know	*The 'Remember' group* did not know how/didn't know how did not know what/didn't know what do not know how/ don't know how does not know how don't know what
	9.	*The 'Tell' group* asked [somebody] call [somebody]	*The 'Tell' group* advice/advised [somebody] asked [somebody] called [somebody] tells [somebody]
	10.	*The 'Try' group* tried/try/trying	*The 'Try' group* tried/try/trying
	11.		*The 'Wait' group* cannot wait
	12.		The 'Say' group suggested
	13.	The 'Other verb in phase' group help	The 'Other verb in phase' group help
	14.		Phrasal modal had to
Adjective and adverb groups	1.		*The 'Able' group* able
	2.		*The 'Adequate' group* ready

	3.	*The 'Astonished' group* shocked, very shocked	*The 'Astonished' group* shocked, very shocked surprised
	4.	*The 'Delighted' group* good	*The 'Delighted' group* excited very happy
	5.		*The 'Willing' group* sincere
	6.		*The 'Easy' and 'Difficult' group* It [be] hard
	7.		*The 'Selfish' and 'Dangerous' group* It [be] to dangerous
Others	1.	about	in order

It must be pointed out that the set of meaningful units of language use to be discussed here is, following the systematic treatment of meaning and usage involving verbs, nouns and adjectives in Francis et al. (1996, 1998), divided into smaller groups according to their basic meaning. Collectively they are, as noted in Chapter 5, known as 'meaning groups' (Francis et al., 1996, p. ix). For example, instances of units in the 'Promise' group are concerned with being committed to a future action. This includes:

- talking or writing about a future action e.g. *agree, promise*
- thinking about a future action e.g. *choose, decide*
- doing something about a future action e.g. *arrange, prepare*

(Francis et al., 1996, p. 91)

Examples falling into and occurring with this meaning group, as attested in the *Bank of English* at the time on which Francis et al. (1996) was based, are: *agree, arrange, choose, consent, contract, decide, determine, elect, go out, fix, guarantee, intend, be looking, mean, offer, opt, plan, set out, pledge, plot, prepare, promise, propose, resolve, scheme, swear, threaten, undertake, volunteer, vote,* and *vow*.

Returning to our discussion of how the developing language users might meaningfully engage in the construction of their narratives, we find from Table 5.5 that they started off, at Time 1, by making use of eight verb-oriented meaning groups and two adjective-oriented meaning groups, as follows:

1. Verb-oriented meaning groups

(a) The ‘Help’ group

(b) The ‘Hope’ group

(c) The ‘Like’ group,

(d) The ‘Promise’ group

(e) The ‘Remember’ group

(f) The ‘Tell’ group

(g) The ‘Try’ group

(h) A meaning group indicating ‘Other verb in phase’

2. Adjective-oriented meaning groups

(a) The ‘Astonished’ group

(b) The ‘Delighted’ group

As we turn to the narration constructed at Time 4, we find a number of notable changes made by the developing language users. First, there is a recurring instance of a new, innovative noun-oriented meaning group, the ‘Thought’ group (i.e., *a good idea*). Second, from the initial eight verb-oriented meaning groups observed at Time 1, the developing language users are now extending the number to thirteen, making use of these meaning groups in addition to one recurring phrasal modal (i.e., *had to*). Third, the number of adjective-oriented meaning groups has substantially changed too, from two to seven. These observations are summarized below:

1. Noun-oriented meaning group

The ‘Thought’ group

2. Verb-oriented meaning groups

(a) The ‘Begin’ group

(b) The ‘Discover’ group

(c) The ‘Help’ group

(d) The ‘Hope’ group

(e) The ‘Like’ group

(f) The ‘Manage’ group

(g) The ‘Promise’ group
(h) The ‘Remember’ group
(i) The ‘Tell’ group
(j) The ‘Try’ group
(k) The ‘Wait’ group
(l) The ‘Say’ group
(m) A meaning group indicating ‘Other verb in phase’
(o) Phrasal modal

3. Adjective-oriented meaning groups
(a) The ‘Able’ group
(b) The ‘Adequate’ group
(c) The ‘Astonished’ group
(d) The ‘Delighted’ group
(e) The ‘Willing’ group
(f) The ‘Easy’ and ‘Difficult’ group
(g) The ‘Selfish’ and ‘Dangerous’ group

Taking all these meaning groups into consideration, we find the narrative world projected at Time 1 to be one made up of a series of activities and events. There are attempts by characters helping someone (i.e., *help, helped [somebody]*; *tried, try, trying*), with relevant communication events taking place (i.e., *asked [somebody], call [somebody]*). The characters have been depicted as people concerned with doing something about a future action (i.e., *planned*) and with knowing or not knowing what or how to do something in critical situations (i.e., *did not know how, didn’t know how, do not know how, don’t know how, does not know how, doesn’t know how, know what, know, didn’t know*). They are individuals undergoing situations that are perceived to be shocking (*shocked, very shocked*) or pleasant (i.e., *good*, as in *The girl's parent are good to hear that their daughter had safe*) in life. They are also individuals with attitudes towards a future action or event (i.e., *do not want, want, wanted, wants*) or something (i.e., *like*).

That tells a great deal about what is going on in the narrative world depicted at Time 1: the details here complement well those revealed by the earlier analysis of emergent units

surrounding *that* in which only the existence of physical objects and limited communication events is related; that would have been a rather worrying state of affairs with seemingly prevalent impersonal modes of interaction if it had been true, even in a fictional world. As we have just seen, however, this is (fortunately) not the case.

How does the imagined world at Time 4 look like? Some aspects are found to remain generally the same. For example, there are similar attempts by characters helping someone (i.e., *help, helped [somebody]*; *tried, try, trying*) and the characters have been depicted as people concerned with knowing or not knowing what or how to do something in critical situations (i.e., *did not know how, didn't know how, did not know what, didn't know what, do not know how, don't know how, does not know how, don't know what*). The characters have also been described with attitudes towards a future action or event (i.e., *want, wanted, wants*).

As can be expected, new details are at the same time added to the newly constructed narrative world. For instance, the characters are now seen to be involved in starting some action (i.e., *started*) and concerned with finding out how to do something (i.e., *learn how*), with descriptions of how something usually difficult has been successfully accomplished (i.e., *manage, managed*), as illustrated in the following examples:

(11) He *managed* to save Balqis from drowning. (011d)
Luqman *managed* to find a long rope near a verdant jungle and quickly tied it to a shady tree …. (038d)
However, he tries again and several seconds later, he *manage* to pull Aini out of the river. (091d)

Also, not only are the characters now shown to like something (i.e., *like*), they are also described as having a love for something (i.e., *love, loves*). In addition to being concerned with doing something about a future action (i.e., *plan, planned, planning*), their commitment to a future action has extended to making promises and decision making (i.e., *decide, decided, promise, promised*) and waiting to do something (i.e., *cannot wait*). Also, communication events that are taking place are not limited only to asking and calling (i.e., *asked [somebody], called [somebody]*) but also include those of telling, advice giving and suggestion making (i.e., *advice or advised [somebody], tells [somebody], suggested*):

(12) The doctor *advised* Lucy to rest for a few days. (022d)
Nisa *tells* Ana to calm down and relax. (121d)
One day, Leong *suggested* to Amri and Chong to go for fishing at the nearest river. (082d)

The characters have been described as individuals experiencing situations that are perceived to be surprising (i.e., *surprised*) as well as those perceived at Time 1 to be shocking (i.e., *shocked, very shocked*). Further, they are individuals not only considered to be able or adequate to perform an action (i.e., *able, ready*) but also willing to offer something or help somebody (i.e., *sincere*). Perhaps most important of all, the characters have been projected as individuals with feelings (i.e., *excited, very happy*) and the narrative world as a place with a series of challenges and dangers (i.e., *it [be] hard, it [be] to dangerous*):

(13) *It is hard* to swim and save someones with shirt. (082d)
Lina had advised her do not pluck the flower because *it was to dangerous* to get them. (118d)
I said that *it is to dangerous* to get the flowers, because the river bank was very slippery. (123d)

This is a long list of observations about conceptual and perceptual changes that have taken place in the construction of narrative discourse as time goes by. The broader set of findings obtained than those in the previous study on *that* suggests the richness of the meaning-making resource at the disposal of the second school students considered in this research, and illustrates the dynamics of meaning-making in the developing language. Essentially a convergence of evidence is witnessed arising from both the present and the previous study, showing how the developing language users are expanding their meaning-making capacities over time.

There is one final note to make. The 'narrative world' thus far discussed is by no means singly constructed by an individual; the (changing) narrative discourse is jointly constructed by the group of developing language users considered in this thesis. In a corpus study, the corpus of texts is treated as if development were a collective effort. What a corpus study (of this type) does not focus on is the individual differences between the writers. It is possible, for example, that what happens is not that all writers develop meaning resources but that writers

become more different from each other, as is the case suggested by the findings of another study reported in this thesis, to be discussed in Section 9.3 of this chapter.

The notion of meaning making

I have used the term 'meaning making' to discuss and explain the series of findings in this section. It might be necessary at this point to clarify what I mean by the term.

The idea of language as a resource for making meaning is certainly not new and is perhaps most fully articulated in a view of language associated with M. A. K. Halliday's systemic functional linguistics (SFL). It is based on the belief that language exists to fulfill the communicative needs of its users (e.g., Halliday, 1994; see also recent attempts documented in two special journal issues published in *Linguistics and Education* (2009) and *Journal of Second Language Writing* (2013), guest-edited by Byrnes, which are devoted to the exploitation of this theoretical perspective to inform the study of instructed foreign language development, and the study of writing as meaning-making and the teaching of writing as teaching to mean, respectively).

That language is used to make meaning is clearly shared in this thesis. Instead of framing the analysis and approaching the data with a top-down perspective using an SFL framework, however, this thesis begins the study of the developing language inductively, using single words as the departing point to investigate language in use. What is observed in the process is, surprisingly, a set of findings converging with the central tenet of SFL: that language is a meaning-making resource. The little insights obtained in this research might contribute to the current debate on the relationship between SFL and corpus linguistics from a meaning-orientation perspective (see, e.g., the collection of papers in Thompson & Hunston, 2006; and Hunston, 2013).

Specifically, the research reported here has approached the developing language without making a distinction between form and meaning (Sinclair, 1991). The earlier discussion on how the developing language users are changing their meaning-making resources over time based on the analysis of emergent lexis and its classification into meaning groups might make clear the approach adopted. Essentially, 'all words and all forms are inhabited by intentions' (Bakhtin, 1981, p. 293).

9.2.3.4 Implications of viewing the developing language as a dynamic meaning-making resource

The last point made by Bakhtin is important. It is based on a formulation of dialogism which posits that at any moment of language use, words are observed, on the one hand, to be carrying historical traces of others' intentions. The use of these words thus contributes to the making up of what we call 'shared meanings', which is a reproduction of 'common' knowledge. On the other hand, the language user must necessarily speak or write to the context before them, bringing in their own voice and intentions constructed over time and discursively shaped by their specific socio-cultural and historical background of language use and contact. This is a creative process involving the language user's own contribution to the discourse.

Depending on context of use, the text produced, spoken or written, becomes then a site of tension with potential conflicting worldviews and incompatible ideological orientations. In the classroom, for instance, those instances that conform to the conventional use of language become almost instantly part of a valued text or language production event; those which are not might be considered 'not clear', 'erroneous' or even, 'nonstandard'. The discussion presented in Chapter 1 on the need for an artificially constructed distinction between conventional and innovative language use is itself a revealing example of how this ideological tension oriented towards a normative frame of mind is endemic even in our very own discipline.

The important point to make here is that every instance of language use by the developing language user is meaningful. It is an expressive production by the developing language user that is no less meaningful than any (other) conventional instance of use. A view that prioritizes a strict adherence to the use of linguistic features and structures from the so-called 'standard' or 'educated' language would only restrict oneself to one's own values and interpretive schema. The argument against such a view is not so much that it is wrong; rather, it is missing the point that each form of human language is valid in its own terms and that it is functional within its user group. This 'valid and functional' perspective applies to what I have called in this thesis the developing language.

To paraphrase the philosopher and sociologist of science Bruno Latour (1988), the bizarre idea that the standard variety of language might be fully transparent, intelligible and systematic in its use is a mirror image of the other no less bizarre idea that 'learner language' might be fully transparent, intelligible and systematic in its use. The challenge then is to recognize that when there are difficulties in making sense of the developing language user's innovative productions within a traditional or conventional perspective, our task is not to dismiss the productions but rather to question that perspective and adjust it to accommodate the new evidence (cf. Sinclair, 1991).

The implications of considering the developing language as a dynamic meaning-making resource are daunting (cf. Sinclair, 2004) on a number of grounds. Two of them particularly stand out and they concern a wider conception of the goals and pedagogical practices we use in schools and a rethinking of the notions of 'language' and 'language learning'.

9.2.3.4.1 'The error-free classroom'

First, if we accept that the developing language user is making no errors but only meaning, the role of the language teacher then is not so much about error correction, as is still the case in many classrooms today. The new role, as it were, is about making the student *critically* aware that first of all, the 'errors' the student makes are no real errors: every word and every utterance or sentence are productively meaningful – at least to the student, if not to the teacher or the classmate as well.

To discard and remove the whole idea of errors from the lives of students does not make the language learning task any simpler, however. In fact, the demands placed on them may be increasing. Mature language user worlds are becoming more complex (and unfortunately sometimes self-justified); with this increased complexity comes the need for language students to explore, respond and make their own choices. Here is where a socially responsive pedagogy comes into play. One way might be to draw insights from what de Certeau (1984) has termed 'appropriation': the student has a *conscious* choice of whether or not to *appropriate* their language use according to the socially preferred norm of practice, and the goals of pedagogy are to engage them in a critical awareness dialogue about the use of different forms of language and the relevant implications each entails.

The view that 'there is no right or wrong, but there is appropriate and inappropriate' is a fairly old one, of course. In this context, however, what is emphasized is an agentive view of 'self-realization from within' leading to self-directed actions and responses, as opposed to an externally imposed notion of 'appropriateness'. This involves the student's making choices and decisions as to whether or not to *adjust* themselves according to the social life of cultural practices.

The difference in this pedagogical orientation entails a shift in thinking (and in practice) from a traditionally held view of the existence of 'slow' or even 'dull' learners repeatedly making the same errors (while constructing new ones), to a view that sees the student as a human agent capable of actively making sense of how the social world in which they live 'works' and capable of making their own informed choices of whether or not to conform to the social norm of practice, bearing in mind the relevant, potential consequences.

9.2.3.4.2 The one-language hypothesis

The second implication concerns how language and language learning might be reconceptualized. Because meaning making, as argued here, is seen as the underlying, guiding force of language use and development, learning more than one 'language' fundamentally means learning a larger set of surface signs, forms, expressions, and sounds as well as writing systems, but no more than that. The same developmental, dynamic meaning-making resource is at work. In other words, it is likely to be the case that there is only one language we all speak at the moment around the world, but because of, again, the various (unknown or forgotten) historical reasons, some social preconceptions have been formed of the existence of 'English', 'French' or 'Mandarin' variously spoken by different populations in different geographical locations on the planet. This seemingly absurd claim is increasingly supported by recent thinking and research findings from the field of bilingualism and multilingualism, where terms like multicompetence (e.g., Cook, 2008) are enjoying a wider acceptance and usage nowadays.

These implications will be discussed further in the next chapter.

To sum up, in addressing Research Question 1:

> How does the use of three selected closed-class words among the developing language users change over time and what does this tell us not only about those words but also about other aspects of the developing language?

we find that the function word *that* changes its relative frequency over time, *to* remains as the second highest frequency word throughout the four points of development, and *of* shows no consistency in frequency as time goes by. This information about relative frequency is, however, not corresponding in any meaningful way to observations of the usage of the relevant function words. The changing patterns of use of the words require a close study of the individual words with their co-text and this, in turn, reveals the changing regularities of association between lexis and grammar in the developing language. At the same time an expansion of linguistic resources is observed, a reduction in the number of patterns of use is witnessed. A further, and most important, observation is that language development is about the construction of a dynamic meaning-making resource. Two implications arising from such a view have been briefly considered.

9.3 Research Question 2

(a) As ELUs repeat the narrative task over two years, do they write more or less?
(b) Does the structure of the narrative remain the same or does it change? If it changes, in what way does it change?

This question, as can be seen, consists of two parts. The first part concerns the changing text length that might take place over a period of two years. The second, more qualitative in nature, asks whether the structure of the narrative would change over time and in what way it might change. In this section, I begin by considering the first part of the question.

9.3.1 Do ELUs write more, or less, over time?

The study reported in Chapter 7 has been conducted to answer this question. It concerns the issue of whether there is an increase or decrease in text length as time goes by based on the production of the number of words in the texts written by the ELUs considered in the study. The results show that there is a general trend towards greater production over time. Most texts

composed in 2007 by the developing language users, for example, fall within the range of 90-250 words. This accounts for 112 or 90 per cent of the total narratives produced at that time. This changes, however, in 2009 in which the same group of ELUs is found to produce a large proportion of their texts within the range of 151-350 words: 105 or per cent of the total texts are of this length.

Over time 109 ELUs or 88 per cent of the whole group in the study have been found to produce longer texts. The increase in text length is between 3.6 per cent up and 265.3 per cent, with approximately one fifth of these longer texts in 2009 found to be at least twice as long as those written in 2007. This coincides well with one aspect of findings from the previous corpus studies reported in this thesis that second language development is, in part, about the expansion of one's linguistic repertoire as time passes by. It should be noted though that the study of a corpus of ELU texts, as in the present case, is essentially a measure of group performance by the developing language users, and that individual ELUs can be expected to change at different rates, deviant from the average or idealization of the overall group performance.

Indeed a major finding from the study is that the developing language users start off as more or less a homogenous group and become much more differentiated as time goes by. The 124 ELUs have been found to display a greater variation in terms of language production when composing their narratives in 2009 than in 2007. As shown in Chapter 7, comparisons using the paired-samples t-test found a statistical difference between the two sets of narratives written in 2007 and in 2009, respectively (mean difference = -83.94, 95% CI = -98.78, -69.11, $p < .0001$). In other words, second language development in instructional settings appears to be a process whereby differences in developmental patterns between and among individual ELUs become more pronounced over time.

How these changes take place as evidenced in the narratives composed by individual ELUs are considered in the next section.

9.3.2 Does the structure of the narrative remain the same or does it change? If it changes, in what way does it change?

The qualitative study of narrative texts written by two ELUs, as reported in Chapter 8, has been designed to address this question. Using a matrix analysis initially developed by Pike

(1981) and later revised by Hoey (2001) to investigate the structure of texts, the study aims to capture changes that take place in the organization of narrative texts, from those written in May 2007 to those in June 2009, by the two developing language users: ELU 094 and ELU 054.

From the texts considered, it has been found that both ELUs are changing their ways of presenting their narratives over time. The first narrative written in 2007 by ELU 094, for example, begins by introducing the three male characters as illustrated in the picture prompt, followed by two female characters. The description goes on in the order that one of the female characters has slipped and fallen into the river, her friend shouts for help, and then the three boys come to the drowning girl's rescue. What is observed from this account is an orderly story, simply told with the sequence of scenes following a linear progression of narration. There are no intricacies or complications involved and the characters just go and do things in the narrative world.

The second text by ELU 094, written in 2009, exhibits a number of different ways in the construction of the narrative. For example, the matrix analysis of this piece of narrative shows how repetition of central units of ideas, or what was termed 'redundancy' (following Hoey, 2001) in Chapter 8, has been carefully applied by the ELU to achieve a sense of coherence for the narration. Key ideas of two female characters plucking flowers and later the drowning female character struggling in the water, for instance, are expressed more than once.

Another important observation is the ELU's depiction of the characters' thoughts in the second narrative. As noted in Chapter 8, the male characters have been projected as thinking and feeling individuals, with their emotion or mental processes of 'worry' expressed and their contemplation of whether or not to help the drowning character noted in the story. This contrasts sharply from the narration presented in the first, shorter text written in 2007 in which what the characters thought about or how they felt at some critical moment of the story is unknown to the reader.

What seems to distinguish the second narrative most from the first, however, is the ELU's changing attention to relating how different characters in the story, as the narrative unfolds, interact with one another at a given point in time. We have seen that the matrix analysis of the first text reveals a direct, linear progression of narration, with the action and response of each

character described one at a time. The analysis of the second text, on the other hand, shows simultaneous attention by the ELU to, for example, three different characters in the resolution part of the narrative. In this instance, different acts by different characters are related to take place around the same time, with a description of the complex life-saving act involving the victim, a heroic boy who is getting her out of the pond and the victim's friend who is performing cardiopulmonary resuscitation to save her:

> Naquib … ran and jumped as fast as lightning into the water. He tried to get Alia's body and brought her to the land. Alia's condition was not stable yet, she fainted. Aina, who was joined PBSM club, helped Alia with CPR treatment. (Text 8.2)

This aspect of expanded attention to several characters at once is absent from the first, straightforward chronological narrative.

Turning now to the next developing language user considered in the study, we find that changes are observable over time too in the ways ELU 054 structures his narratives. His first text written in 2007 is largely a straightforward narrative. The characters and events in the story are first introduced following a linear order of narration but unlike the first text by ELU 094 discussed above, there is some redundancy observed towards the end of the narrative: the idea of drowning, for example, is expressed twice in his first text while the act of saving someone is mentioned three times. In other words, the narrative begins and progresses in a strictly orderly manner but ends with some redundancy to reinforce central ideas of the narrative.

The second text written by ELU 054 in 2009 shares this feature of redundancy. The act of a girl's shouting for help is, for example, noted twice; so is the three male characters' attempt to find out what has happened to the girl. There is no so much emphasis on the drowning event indicated in this second narrative, but a reference to the act of saving the girl's life is, as in the first narrative, made three times. Both first and second texts then can be said to make use of redundancy as a way of narration, with the second narrative displaying or exploiting this feature relatively more extensively.

A notable change observed in the second text, when compared to the first, is how the narration has been complicated through the introduction of a greater number of characters. While the first narrative is concerned with five main characters (three male, two female) as

shown in the picture prompt, the second narrative further introduces imagined characters which are not illustrated in the prompt: a nurse (with an ambulance) and the parents of the drowning girl. In this latter narrative, the 'new' characters are contacted by one of the male characters after he has learnt that the girl has fallen into the river, and the parents thank the boys for their heroic act.

One final but important change witnessed in the second narrative by ELU 054 is a distinctive feature shared in the second narrative by ELU 094 considered above: the narrator's simultaneous attention to more than one character at the same time. In this case, the feature is illustrated in the following lines:

> … Ali jump into the river after we knew that was happening there and I went to the nearest public phone to call the ambulance and her parents. (Text 8.4)

This is an aspect of narration absent in the text written in 2007 by either ELU. Arguably, this newly introduced and constructed feature, together with those mentioned above, forms part of the rich resources the two ELUs have developed, and signifies their active engagement in the meaning-making process: on the one hand, there are narrating practices observed that are unique to each individual (e.g., ELU 094 begins to take into account the characters' thoughts and emotions in the second text while ELU 054 enriches his latter narrative with more characters); on the other hand, both ELUs have been found to pay simultaneous attention to what different characters are doing in their respective narrative and share this 'expanded view' of storytelling.

To some extent, these observations reflect two modes of meaning making highlighted by Blum-Kulka (2005) in her work on young children's conversational storytelling. One mode, as recontextualized here, refers to a socio-cultural arena within which the developing language user negotiates meanings and relationships unique to their immediate, local culture. The other mode refers to a discursive arena within which the concern is with the development of discursive skills (in this case, narrative-discursive skills), as stepping stones for 'adult-like uses of language and for gaining membership of adult cultures' (Blum-Kulka, 2005, p. 151; see also Bakhtin, 1981, on dialogism).

To summarize, both ELU 094 and ELU 054 show changes over time in the way they construct their narratives. There are similarities as well as differences observed in the construction of their narratives. ELU 094, for example, has been observed to begin with a straightforward narrative that is humbly told. As time passes by, she enriches her narrative by bringing a human touch to the task of relating the characters' states of mind and feeling in her story and by making good use of redundancy to complete her task. ELU 054, on the other hand, starts off with a linear progression of narration and ends his first narrative by introducing some feature of redundancy. Over time, he makes more extensive use of redundancy in his narrative and creates a greater number of characters in his story. Alongside richer characterizations of the narratives crafted in 2009, a remarkable quality shared by these two narratives is an expanded view on the part of the narrator, or change of perspective, which allows for a consideration of more than one character at a given time. Taken together, all this, it is argued, points to the developing language user as a social being who is actively engaged in the development of 'the discursive skills of storytelling while making sense of the world' (Blum-Kulka, 2005, p. 169).

9.4 Evaluation

In this chapter two research questions have been considered using the results of a series of five studies as reported in this thesis. These studies, in turn, have been conducted using two main approaches to language and (written) discourse analysis. The first three lexis-driven studies are based on a corpus approach while the analysis of individual narratives is based on a text-based approach known as a matrix perspective. In this section I consider the relative merits of each approach.

9.4.1 Corpus approaches to researching the developing language

Studying the developing language from a lexis perspective essentially entails, and reflects, an inductive approach. Also known as a corpus-driven approach (Tognini-Bonelli, 2001) or a word-based method (Hunston, 2002), as noted in Chapter 2, an inductive enquiry has in this thesis enabled an exploration of the raw language data with minimal interference or bias from the researcher. This does not, however, imply a naïve assumption that the researcher is working within a theoretically free position. Neither does it suggest that the idea that this contrasts sharply from an approach that prioritizes a reliance on linguistic concepts and categories that exist in pre-corpus days is unproblematic (see, e.g., McEnery & Hardie, 2012; see also the papers in Worlock Pope, 2010).

Rather, it is, as Hunston (2002, pp. 92-95) points out, part of a cycle involving one's theoretical orientation and practice. As an illustration, before conducting the studies reported in this thesis, I have collected and assembled a longitudinal corpus of language samples by developing language users over two years. As far as my theoretical inclination is concerned, I share the view that it is "unnecessary to assume any 'facts' prior to statement" (Firth, 1957, p. 199) and that it is important to approach real data with an open mind (Sinclair, 1991, 2004; see also McCarthy, 1998).

This leads me to the decision of keeping the longitudinal corpus as it is, working with this raw (i.e., unannotated) corpus, and then perhaps unsurprisingly, using lexis as the starting point of my linguistic investigation. All this in turn has enabled me to uncover some interesting aspects of the underlying workings of the developing language (see Chapters 4-6) and led to a refinement of my understanding of the nature and process of second language development. My own perspective on the notion of 'language students' has changed accordingly, from initially seeing them as mere learners of language (as opposed to learners of technology, for example) to now viewing them as dynamic meaning makers (see also Chapter 10 for a consideration of the contribution of the study of a longitudinal corpus to Applied Linguistics).

It is important to note that this orientation and practice in approaching language turns out to work particularly well with 'learner data'. This is especially so when one's viewpoint is to treat it in its own terms, unencumbered by preconceptions based on the analysis or evidence of an external norm or point of reference, which is the position adopted in this thesis. Research that traditionally deals with 'learner language' has revealed how idiosyncratic this body of language is, and a lexis-oriented approach allows for an analysis of the developing language that is in keeping with a respectful view of the language.

Concordance lines from the raw corpus not only present evidence and confirm the common-sense intuition that innovative features as well as conventional ones are a common aspect of the developing language; more importantly, they demonstrate that the ways in which innovative uses of language manifest themselves are many and varied. Working with a longitudinal corpus further sheds light on the changing patterns of occurrence of these diverse uses of innovative language over time, in interaction with those features of conventional language. As shown in the studies reported in Chapters 5 and 6, all this illustrates the dynamic,

organic nature of language development, where there is no single developmental pattern that can account for how specific linguistic features grow in the living language, but a complex series of them (see Chapter 10 for further details).

A note of caution is in order before we turn to the discussion on the merits of using a matrix approach to analyze texts. It is important to remember that in all corpus studies, the insights offered by the analysis of a corpus are to a large extent as good as the corpus itself. The series of lexis-oriented studies reported here is no exception. The studies have been conducted based on a longitudinal corpus; interpretations of the results might therefore best be made in relation to the nature and scope of the corpus.

9.4.2 A matrix approach to researching narrative writing development

While a lexis-driven corpus approach highlights typical patterns of language use, conventional or otherwise, in a corpus, a matrix approach, as also exploited in this thesis, allows for a more in-depth analysis of individual texts from the same corpus.

The matrix analysis is, like the lexically oriented approach, part of an inductive methodology that is in line with an overall commitment in this thesis to treating the developing language in its own right. As we have seen, there are no a priori theoretical expectations imposed on the narratives studied.

Through the matrix analysis, changes in structuring the narrative by individual ELUs have been observed over time. Details captured in such changes include the introduction of (or a more extensive use of) redundancy and change of perspective in narration. These are features which are not readily observable through a corpus study and suggest the strength of conducting a matrix analysis of texts. While complementing the corpus approach adopted in this thesis, it should be pointed out at the same time that given the nature of the analysis, this approach to text analysis has allowed for the study of a limited number of texts by individual ELUs.

It is perhaps also important to recall that the four narratives considered are written by two ELUs who have shown an increase in text length over time, and these are two out of 88 per cent of the whole group of 124 ELUs who have been found to produce longer texts in 2009. Narratives by the remaining 22 per cent of ELUs who have shown a reduction in text length

over time have, for obvious reasons of space, been excluded from analysis. While space constraints are a real concern, we may ask whether the same matrix analysis works equally well in revealing changes in the narrative structure of texts by those ELUs who write shorter narratives over time.

A quick survey of the relevant texts (i.e., texts written in 2009 that are shorter than those first written in 2007 by the respective ELUs) suggests that this is possible to the extent that the narratives are written in a comparable manner in terms of consistency with the prompt given. In other words, provided the texts are both sufficiently relevant to and reflective of what is depicted in the picture prompt, a matrix study may prove to be not as useful. A case in point is the texts written by ELU 007 (the first text written in 2007 has been presented as Text 9.1 while the second text written in 2009 as Text 9.2).

Text 9.1 First text by ELU 007 (written in 2007)

Atiqah live in Kampung Pandan. She loves to collect beautiful flowers such as orchid, hibiscus and carnatio. She always picking flowers with her friend, Hidayah. They have been collecting flowers since childhood.

One day, Atiqah asked Hidayah to join her picking flowers near the rivershore in Kampung Pandan. At first, Hidayah did not wanted to go there because the river was to deep. Both of them can't swim. But she can't let her did it by herself. Who knows what kind of danger might happen. An advice can't stop her because she was very stubborn.

While Atiqah was picking flowers. Suddenly, she fell into the river. Hidayah can't swim to save he, so she called for help. Fortunately, there were some boys fishing beside them. One of the boys jumped into the river and saved Atiqah. The girls were very grateful to the boys.

Atiqah went home with her wet clothes. Her parents was shocked when they saw her. Her father asked her about the wet clothes. They were very angry when she told them the whole story and she was grounded for two weeks. From this day, Atiqah will always listen to others advice especially her parents and friends.

Text 9.2 Second text by ELU 007 (written in 2009)

I have a penfriend named Ali. We have been sending letters to each other since last year. We exchange stories, experience, and many more. The story that is the most interesting to me is how he got a girlfriend.

He told me that he got the girl when he was going fishing on 14th October 2008. The girl he mentioned fell into a river when she picking some flowers. Ali and his friends heard their scream for help and quickly saved the girl. They manage to rescue her and send her to the hospital.

As can be seen, the first text (Text 9.1) is a direct response to the details shown in the picture task (see Appendix). On the other hand, the second and much shorter text, which consists of two paragraphs, may be considered partially relevant. The part considered relevant for a matrix analysis is the second paragraph of the text: it is a narrative account presented on the different characters and what they are doing (one girl has fallen into the river and the boys come to her rescue), as illustrated in the picture prompt.

A matrix study may be conducted for both texts, but will almost certainly suggest that the second text written in 2009 is 'underdeveloped' – it is in need of details. Only the second paragraph of the text, as noted above, is relevant and thus readily analyzable. From a narrative-as-performance perspective, however, this 2009 text signifies an emergent element of contextualization that is missing from the first text written in 2007. The element in question is, perhaps ironically, reflected in the apparent irrelevant first paragraph of the 2009 text:

> I have a penfriend named Ali. We have been sending letters to each other since last year. We exchange stories, experience, and many more. The story that is the most interesting to me is how he got a girlfriend. (Text 9.2)

That is, what is observed here is the use of meta-narration, "those devices which comment upon the narrator, the narrating, and the narrative both as message and as code" (Babcock, 1977, p. 67). This is a host of devices which have been noted to be often marginalized or overlooked for their assumed irrelevance to the narrated events, leading narrative scholars such as Georges (1981) to ask, over three decades ago, 'Do narrators really digress?' in a paper considering the notion of 'audience' in narrating.

How the use of meta-narration meaningfully contributes to the construction of a narrative has also been pointed out by Bauman and Briggs in their oft-quoted paper:

> … meta-narrative devices index not only features of the ongoing social interaction but also the structure and significance of the narrative and the way it is linked to other events. For example, Texas storyteller Ed Bell embeds the following metanarrative comment in a story about a giant bee tree: "And I don' t blame y' all if you don't believe me about this tree, because I wouldn't believe it either if I hadn' ta seen it with my own eyes. I don' t know whether I can tell ya how you could believe it or not, but that was a big tree" … such interventions bridge the gap between the narrated event and the storytelling event by reaching out phatically to the audience. (Bauman & Briggs, 1990, p. 69)

The concern for a productive enquiry then is perhaps not so much to do with the analyzable, relevant part of a narrative; on the contrary, it is the seemingly irrelevant part that escapes from the analysis. For a matrix study to be a worthwhile and rewarding pursuit, the question is therefore not simply of text length but of the relation between what is narrated and what is illustrated in the picture prompt and the extent to which this is established. As for features such as meta-narrative devices that are important to be taken into account in a developmental or longitudinal study such as the present one, the researcher might devise a separate analytical tool or simply add to the existing analysis comments that relate to the 'new' finding.

The concerns aside, a matrix approach to researching writing development has yielded some important findings. The different kinds of complexity captured through the study of texts by ELUs who present longer narratives over time are particularly revealing. As discussed in Chapter 8, when there are more ideas introduced to a piece of narrative, the narrative becomes more complex. This is a case of *quantity* leading to complexity. Similarly, when ideas of a different nature are introduced (e.g., narrating expressions of the characters' thoughts and emotions versus describing the characters' actions), complexity is recognized here as a matter of quality. This then becomes a case of *quality* leading to complexity. Both kinds of complexity have been observed in this study.

Further, complexity may take the form of redundancy. For instance, the first text by ELU 094 records events as they are seen from the prompt and describes them in strict sequence. The second text, on the other hand, has been found to make use of more explicit reference ties within and across paragraphs throughout the text. This 'cross-referencing' practice, or redundancy, leads to an observation of emergence of complexity in narrative progression. Finally, different levels of control over the change of perspective also suggest different levels of complexity in the narrative. I have used Hoey's (2001) analogy to show how the first texts written in 2007 by the two ELUs considered in this study represent the work of a narrator who, more or less like the one-handed puppeteer, puts down one character before picking up the other. Over time, however, both ELUs have been observed to make an advance in terms of the level of sophistication in control over the change of perspective in narrative, paying simultaneous attention to more than one character at once.

In sum, both the lexis-oriented corpus approach and the matrix approach are part of a useful inductive methodology for investigating development of language and narrative writing. The

strength of the former is in allowing the researcher to access all instances of a selected word (including such function words as *that*, *to* and *of* as in this thesis) with its accompanying lexical and grammatical patterns in the entire corpus, and in a longitudinal corpus, to identify the changing patterns of use of the word. It presents information as is available in the (raw) corpus, much to the delight of the researcher with an open mind. The findings obtained may then be formulated into statements about how language works in that particular corpus, with insights sometimes generalizable to wider contexts (e.g., Sinclair, 1991).

Taking a corpus approach essentially means the individual texts in the corpus are not given due consideration. This by no means suggests a weakness of the approach; no methodology accounts for all aspects of research needs. The individual texts are considered in this thesis instead through a matrix analysis. Qualitative in orientation, a matrix investigation is part of a broader text-based approach which provides answers to a different set of questions and allows for a relatively in-depth analysis of the structure of a text. In a longitudinal study such as the present one, it reveals how changes take place in texts produced at different points in time. The extent of the usefulness of this approach, however, depends on the extent of how narratives constructed at different times are consistent in representing the details of a prompt.

9.5 Conclusion

This chapter has addressed two research questions initially raised in Chapter 1. A series of related findings based on LoCDeLUNT have been presented and discussed. The analysis of three selected closed-class words (*that*, *to* and *of*) in the longitudinal corpus has illustrated not only how these words regularly work in the company of other lexis and patterns, but also how the developing language represents a dynamic meaning-making resource. The results suggest the need for changing and placing the developing language user at the centre of an agent-oriented view of theorizing, and raise important questions about the nature of language, language learning and pedagogical applied linguistic practice, all of which will be considered further in the next chapter.

The agent-centred view of the developing language user is further supported and reinforced by the results of the next analysis involving all the individual texts written in 2007 and in 2009 by the 124 ELUs in this thesis. Variations in text length between narratives by different ELUs have been found to become more distinct, rather than more common, over time. When a set of individual texts are further considered through a matrix analysis, it becomes

shockingly apparent that there is little room for the developing language user to remain negatively framed as glass half-empty (at the time of writing, for example, a quick look at the table of contents of a 2014 issue of *Studies in Second Language Acquisition*, one of the leading journals in Applied Linguistics, reveals that error correction is the focus of one research article). The texts studied from a matrix perspective show how both common and unique features can be observed across texts written at different points in time by the same ELU as well as across texts written by different ELUs, demonstrating the dynamic nature of the individual ELUs who are at times conforming to conventional, discursive constructions and at times displaying the creative capacity in their ever changing narrative practices, just like any other human narrator.

In this chapter, an evaluation is also presented of the two main approaches adopted in this thesis to researching language and writing development. It must be emphasized that their adoption is not accidental: they reflect my theoretical orientation and practice, and they are well-suited to the task of addressing the different sets of research questions raised in my quest for understanding how language development, including writing development, takes place over time.

The next chapter is devoted to a consideration of two remaining research questions of this thesis. The discussion in both this and the next chapter, I hope, reinforces my argument that the developing language user is, in or outside the classroom, an equally respectable fellow communicator and meaning maker. It is also my hope that it opens a way towards illuminating the agentive nature of human language learning and use.

Chapter 10

Discussion (continued)

10.1 Introduction

This chapter considers the third and the final research questions of the thesis. It begins with an assessment of the extent to which it is possible to research and analyze the developing language in its own terms when tracking changes that take place over time in the process of language development. Next it considers the contributions of the study of a longitudinal corpus to learner corpus research and to our understanding of language development. Finally the chapter draws together relevant insights from the empirical findings of the studies documented in this thesis and from a number of disciplines to propose a pedagogical model that aims to revitalize how practitioners and policymakers go about thinking and acting to promote an approach that redresses the balance between the deficit and alternative models in pedagogical applied linguistics.

10.2 Research Question 3

> Given the ideological standpoint that I approach the task of researching the developing language with respect, for it to be viewed as another form of natural, human language in its own right and not against a native-speaker norm or an external point of reference, is it possible for me, the researcher, to remain true to that standpoint while I am studying the changes in the developing language that take place over time?

This is a challenging question to address. The purpose of having and keeping this research question is more about starting dialogues and discussions than about attempting to offer a 'definite' answer. I present two competing 'meta-perspectives', or what I call 'metaviews' in this chapter, which relate to two possible, different responses to this same question. The complication of the whole issue will become apparent when the second metaview is considered, which is informed by insights from other (related) disciplines. In this latter case, I draw heavily from the work of the educationist Ronald Barnett (2013) and the philosopher and cultural critic Slavoj Žižek (2006). A discussion is then presented on the need for a revision and expansion of conceptual frameworks concerning language use and language learning.

10.2.1 First metaview: A matter of perspective

The first response to the question, with its associated perspective, would be a 'yes'. This is a position that runs counter to the practice of, in the words of Larsen-Freeman (2006, p. 590), 'most researchers in the field'. It was lamented in the 2006 paper that these researchers hold the assumptions that 'there are fixed, homogeneous native and target languages' and that second language development is 'a process of increasing conformity to a uniform target language'. When making the comments, Larsen-Freeman was of course highlighting a complex, dynamic system perspective with which she is now most closely associated. The observations made are still largely true today, with language development often being viewed as 'a transitional state that is (or should be) ever changing *towards* the target [language]' (Ortega, 2009, p. 140, italics original; see also Ortega, 2013).

It should also be pointed out that the position I have adopted of treating the developing language in its own right is certainly not new to the Applied Linguistics community. As reviewed in Chapter 2, the general idea that 'learner language' is an autonomous linguistic system can be traced back to the work of Corder (1967, 1971), Nemser (1971) and Selinker (1972). Further, a number of prominent researchers have over the years called for approaching 'learner language' in its own terms and not against a target language norm (e.g., Huebner, 1979; Bley-Vroman, 1983; Cook, 1992; Klein, 1998; Lakshmanan & Selinker, 2001).

Despite these scholarly efforts, however, it remains an acknowledged challenge in the field to find ways to study 'learner language in its own right, rather than as an imperfect version of the target grammar' (Ortega, 2009, p. 143). In addition, the dominant research practice based on a deficit view has been observed to be directly responsible for the lack of disciplinary impact in SLA as a field (Klein, 1998). All this leads Ortega, in a recent state-of-the-art review article, to ask the following:

> With such distinguished SLA advocates, and with such powerful and increasingly comprehensively articulated arguments, why has SLA had such difficulty breaking away from the straightjacket of the comparative fallacy, the target deviation perspective, and the monolingual native speaker bias? (Ortega, 2013, p. 15)

As noted in Chapter 2, the same state of affairs is also evident in learner corpus research.

The series of studies reported in this thesis is collectively a direct response to the challenge of addressing the developing language as an equally valid form of natural human language, researching it as what it is instead of what it is not (e.g., a deficit version of a target language). Here innovative language use is respected and appreciated as much as conventional language use. The findings based on the approach adopted, together with the relevant implications, have been highly revealing of the nature and process of language learning and use, as summarized in the previous chapter and as will be further considered in the remainder of this chapter. In some instances, novel observations such as the paradox of complexity in the developing language have further been noted (see Section 10.3 below).

What is observed then is a familiar cycle involving one's theoretical position and practice (see Chapter 9): a theoretical commitment to treating the developing language with respect has led to a counting practice and consideration of all the recurring features of the language as equally valid instances of language use, which in turn has led to novel orientations to language use and language development.

The overall discussion here points to two broad, different perceptions of one phenomenon that exist in our field. What is a deficit version of an idealized competence for some researchers is considered by others (myself in this thesis, for example) an equally functional, natural form of human language. There can be no one true reading: it all depends on one's perspective. That is, depending on one's point of view, the developing language can be dressed up either as a site of natural language use or as a site of error production, and there is no deciding between these points of view.

This is a case of the familiar 'two sides of the same coin' argument. The one phenomenon can be viewed in quite different ways, depending on one's frame of mind. The one phenomenon, indeed, bears witness to both 'natural language use' and 'imperfect language performance'. Both natural linguistic instances and deviant features are considered real; that is not to be denied. One's approach to the developing language or to the imperfect 'learner language' and the vantage point each offers are, however, liable to yield quite different accounts.

The observation of the paradox of complexity in the developing language in Chapter 5 is a good case in point. The dominant 'deficit' view would interpret the same finding reported therein as a reduction in errors and an increase in complexity, so that the 'learner language'

would be seen to be becoming more complex as time goes on while the 'errors' would be excluded from a consideration of a dynamic view of complexity. When recurring instances of innovative language use are considered as equally valid instances of patterns, however, we start to note a reduction at the same time in the number of patterns of language use, leading to the observation of an emerging picture of the dynamics of complexity that is simultaneously declining and increasing in the developing language.

10.2.2 Second metaview: The parallax shift

A different metaview in understanding the interaction between a perspective on deviant features (based on a prescriptive, deficit view) and a perspective on natural linguistic occurrences (based on a descriptive view) in the analysis of the developing language may be found in Slavoj Žižek's idea of parallax and the image of the Möbius strip. Here, we encounter

> phenomena which are mutually untranslatable and can be grasped only in a kind of parallax view, constantly shifting perspective between two points between which no synthesis or mediation is possible. Thus there is no rapport between the two levels, no shared space—although they are closely connected, even identical in a way, they are, as it were, on the opposed sides of a Möbius strip. (Žižek, 2006, p. 4).

First productively applied in an educational context in an analysis of issues surrounding the potentials and risks of collaboration in the academia (Barnett, 2013), this parallax view holds that the phenomena are essentially same sides of the same coin.

As Žižek (2006, p. 29) elaborates:

> Every field of "reality" (every "world") is always-already enframed, seen through an invisible frame. The parallax is not symmetrical, composed of two incompatible perspectives on the same X: there is an irreducible asymmetry between the two perspectives, a minimal reflexive twist. We do not have two perspectives, we have a perspective and what eludes it, and the other perspective fills in this void of what we could not see from the first perspective.

In this view, the natural linguistic instances and the errors 'hang together': they are dependent on each other. This mutual dependence is part of the total 'fact' of the developing language. It is, however, very difficult, if not impossible, to hold both in view at once; at any one moment, we are drawn either to focusing on the errors or on the natural linguistic instances.

A similar view is offered by Deleuze (1994) in a discussion of the writing process:

> We write only at the frontiers of our knowledge, at the border which separates our knowledge from our ignorance and transforms the one into the other. (Deleuze, 1994, p. xxi)

That is, a similar 're-entering' process is highlighted here in which our knowledge and ignorance are constantly negotiated.

Where is this set of reflections – and perhaps seemingly arcane reflections – taking us? Namely this: that in considering the developing language, we have an example of the extraordinary complexity that makes up for the nature and process of language development. They are testimony to layers and facets of reality, which run into each other.

Let us recall the Möbius strip. One can trace one's finger from a point on one side along its entire length, only to find that after a circuit, one finds oneself on the opposite side from one's starting point. One side runs into the other.

Another point equally relevant to our discussion in relation to the parallax view is the idea that things do not just appear but rather, 'appear to appear'. This may be exemplified by the analogy of a frame for a painting:

> The frame of the painting in front of us is not its true frame; there is another, invisible, frame, the frame implied by the structure of the painting, the frame that enframes our perception of the painting, and these two frames by definition never overlap—an invisible gap separates them. … The frame is always-already redoubled: the frame within "reality" is always linked to another frame enframing "reality" itself. Once introduced, the gap between reality and appearance is thus immediately complicated, reflected-into-itself: once we get a glimpse, through the Frame, of the Other Dimension, reality itself turns into appearance. In other words, things do not simply appear, they appear to appear. (Žižek, 2006, p. 29)

The point made is an important one:

> … once things (start to) appear, they not only appear as what they are not, creating an illusion; they can also appear to just appear, concealing the fact that they are what they appear to be. (Žižek, 2006, p. 30)

Note that this 'appear-to-appear' argument may, in one sense, be taken to suggest a view that is against a 'deficit' model of conceptualizing the developing language: certain features of the

developing language 'not only appear as what they are not [i.e., as signs of stigmatized 'learner language'], creating an illusion', they also 'appear to just appear, concealing the fact that they are what they appear to be'. In fact, this is a recurring argument of this thesis: that for various social, cultural and historical reasons, naturally-occurring linguistic signs and symbols expressed or made by different groups of language users ('learners' versus 'native speakers') have over the years been interpreted and valued differentially.

The main insight to be gained from this parallax view is, however, not so much that it supports my argument; in actual fact this metaview supports neither the position I hold nor the position based on a deficit model. As Žižek illustrates through a Buddhist orientation as a way of closing his 400-page book,

> A wonderfully ambiguous indicator of our present ideological predicament is *Sandcastles: Buddhism and Global Finance,* a documentary by Alexander Oey (2005) with commentaries from the economist Arnoud Boot, the sociologist Saskia Sassen, and the Tibetan Buddhist teacher Dzongzar Khyentse Rinpoche. Sassen and Boot discuss the gigantic scope, power, and social and economic effects of global finance: capital markets, now valued at an estimated $83 trillion, exist within a system based purely on self-interest, in which herd behavior, often based on rumor, can inflate or destroy the value of companies—or whole economies—in a matter of hours. Khyentse Rinpoche counters them with ruminations about the nature of human perception, illusion, and enlightenment; his philosophico-ethical statement "Release your attachment to something that is not there in reality, but is a perception," is supposed to throw a new light on the mad dance of billion-dollar speculations. Echoing the Buddhist notion that there is no Self, only a stream of continuous perceptions, Sassen comments about global capital: "It's not that there are $83 trillion. It is essentially a continuous set of movements. It disappears and it reappears." (Žižek, 2006, pp. 383-384)

Framing our discussion within this metaview, we find then that the issue is not so much about which position makes more sense or reflects the 'reality' better; rather, the 'new' understanding is that both competing positions are only part of 'a stream of continuous perceptions' or 'a continuous set of movements'. It is the future, or the mind (perhaps a 'distant Self'), who has the last word (see also Hofstadter, 1999).

It is important to recognize that this parallax view, too, is just an idea. It is another way of considering perspectives, one which provides no answer to the research question stated at the

beginning of the chapter but which, I hope, helps to highlight the role of (re)framing and the relevant implications.

10.2.3 Position, foundational assumptions and expanding conceptual options

While I have noted that my position of treating the developing language with respect is only a perspective (which contrasts with another that is based on the dominant, deficit model), it is a perspective that is decidedly more pleasant to live with than the alternative, at least for many developing language users.

As will be discussed in more detail in Section 10.4 below, this position, together with its pedagogical implications, also draws on a number of current and resurgent ideas in the literature of medical sociology (the asset model), management theory (Appreciative Inquiry) and applied linguistics. It places strong emphasis on helping one to understand and discover what could be, rather than on trying to fix what might be (Bushe, 2013): it seeks to develop foundations of positive patterns of growth among developing language users rather than the foundation of negative outcomes (cf. Morgan & Ziglio, 2007).

To state my position more explicitly, it is a set of (constantly renewed) ideas with a reframing of current dominant views of the 'learner', language and the learning process. There are three important, interrelated points to make here regarding the position. First, this position is fundamentally humanistic in orientation. The developing language, for example, has been found in the series of the studies presented in this thesis to be highly complex, dynamic and organic. It is in constant change, consistent with Jespersen's (1909-1949, p. v) view of language noted in Chapter 1 that it should not be thought of as a set of rigid dogmatic precepts, according to which some things are correct and others absolutely wrong, but as something living and evolving under continual fluctuations and undulations, something that is essentially, in one word, human. Regardless of whether it is conventional or innovative language use, the deep human lived experience of using and interacting with language in the personal and social world of the developing language user is valued. Opportunities for engagement in a range of sociocultural situations, dominant or otherwise, are seen as continuously enriching the language user's ways of perceiving, feeling, talking and thinking about reality (e.g., Agar, 1994; Kramsch, 2006).

Second, this position rejects the 'traditional' view which sees language as a body of knowledge to be 'discovered', with superficial dichotomies constructed concealing the dynamic and fluid nature of human language use and development. These dichotomies include, among others, 'accurate' versus 'inaccurate' language use as just indicated above, as well as such constructs as 'learner' and 'native speaker': irrespective of the conventional labels, all language users, 'novice' or 'mature', are essentially *developing language users*. We are all language users in development actively engaging, and being engaged, in the meaning-making process of the world in which we live.

In this position, the notion or the plausibility of using 'emergent language users' or ELUs is also to be questioned and reconsidered: this is a term first introduced in Chapter 1, reflecting my (then) initial idea at an early stage of this research. Readers of the previous chapter (Chapter 9) would no doubt have spotted my use of ELUs in one place and my (now) preferred term of 'developing language users' in another in that chapter. That is the result of having initially set the research questions with the term 'ELUs' in Chapter 1 and then having to, accordingly, answer them in Chapter 9. After being informed by the findings from the study of the longitudinal data and developing an empirical understanding that it is essentially the dynamic process of meaning making that underlines language learning, I find the term 'emergent language user' less helpful than initially thought. This, incidentally, also points to a meaningful professional growth on the part of the researcher in interacting with longitudinal data.

This very term of 'ELU' is likely to mislead one into believing and accepting an equally questionable construct, 'language proficiency', and implies that a linear developmental progression might take place from 'emergent' to 'advanced' language use in language learning. As argued in Chapter 5, the linguistic resources one enjoys at different points in time are inherently complex in different ways throughout the course of language development. The notion of the *developing language user* is therefore most consistent with a position against a deficit framing and is used hereafter to refer to all human communicators who are exploiting language to make meaning in their daily lives. This is also in line with the claim made in Chapter 9 which posits that all people speak and share one language when meaning making is considered the driving force of language use. We keep *developing* our language throughout our lives.

Finally, as should be apparent by now, this position comes with a set of new concepts that are necessary to help to illuminate renewed understandings of some familiar ideas and phenomena, and to reexamine the dominant cultural history and practices in the social life of language use. Alongside the notions of the *developing language user* and the *dynamic meaning maker*, there are four other important terms characteristic and reflective of the position expressed in this thesis. They are *languaging*, *languaculturing*, *supercomplexity* and *simplicity*.

10.2.3.1 Languaging

My understanding and use of *languaging* is both similar to, and at the same time different from, the same term used by a prominent SLA figure, Merrill Swain, and by two noted scholars in the study of modern languages in higher education, Alison Phipps and Mike Gonzalez. Swain uses the term to refer to an understanding of language as a process rather than an object. Languaging to Swain is an activity, a "process of making meaning and shaping knowledge and experience through language" (Swain, 2006, p. 98). The same emphasis on process, meaning making and experience shaping is observed in Phipps and Gonzalez when they discuss languaging:

> Languages are more than skills; they are the medium through which communities of people engage with, make sense of and shape the world. Through language they become active agents in creating their human environment; this process is what we call *languaging*. (Phipps & Gonzalez, 2004, p. 2, italics original)

All this emphasis on process, meaning making and experience shaping is shared in this thesis. In addition, however, languaging, to me, involves flexible language use, encompassing the traditional notion of 'code-switching'. That is, whether an individual 'mixes languages' in their speech or writing (i.e., spontaneously alternating and changing language use between 'two or more languages' in one stretch of discourse), the individual is engaged in languaging. Note too that I have used quotation marks when I discuss 'languages' and 'their mixing', for I hold the one-language hypothesis: that we all speak and share *one* language.

The closest in approach to this broader conceptualization of languaging is found in García (2014). The term he uses though is *translanguaging*:

> … translanguaging is not used to refer to two separate languages or even the shift of one language or code to the other …. Rather, translanguaging is rooted in the principle that bilingual speakers select language features from a repertoire and "soft assemble" their language practices in ways that fit their communicative situations … that is, bilinguals call upon different social *features* in a seamless and complex network of multiple semiotic signs, as they adapt their language practices to suit the immediate task. (García, 2014, p. 3, italics original)

Like García's, my take on languaging is transformative: 'it attempts to wipe out the hierarchy of languaging practices that deem some more valuable than others' (García, 2014, p. 4). Whether it is 'standard English' or 'learner English' (or 'Japanese', 'Malay' or 'Mandarin', for that matter), each is an equally natural form of human language use. Differences in patterns of language use, I argue, must be explained not by focusing on the unique languaging practice of individuals but by considering the larger structure of institutional and social practice. Dominant sociocultural practices set the 'rules of the game' within which convergence of individual practices matters, but this at the same time highlights the fact that what we face is a structural problem, not a hereditary or deficit condition within the individual language user.

Another point relating to languaging concerns the state of the mind of the individual or community. It is hypothesized that it is closely associated with Cook's (2012, p. 1) notion of *multicompetence*: 'the knowledge of more than one language in the same mind or the same community'. Note though that my preferred term in this context is just 'competence' and that it is in a fluid state, in line with the early argument that there is only *one* language and consistent with the conception of languaging. All this, of course, calls for future empirical research findings to support, or refute.

10.2.3.2 Languaculturing

When the notion of languaging is taken seriously, one begins to see that what has been commonly known as language learning (e.g., studying a 'new' language at school, or interacting with other language users and acquiring language informally) is essentially a languaculturing process: the developing language user is constantly, knowingly or unknowingly, being languacultured as well as languaculturing others (e.g., when socializing with other languaging beings).

Whereas languaging signifies the process of language use, languaculturing gives prominence to the intricate interplay between languaging and sociocultural practices of human societies. Based on the conception of 'languaculture' first introduced by the anthropologist Michael Agar in 1994, my use of the term languacultur*ing* places, again, a strong emphasis on understanding the notion as a dynamic process of how one is using and developing one's linguistic-cultural resources. The argument here is that the languaging being is fundamentally a languaculturing being, who is in engagement with moment-by-moment self-renewal.

Similar observations about the interdependence between language and culture have been made in child language acquisition research. In discussing and assessing Jerome Bruner's contribution to this area of study, for example, Shanker and Taylor (2001) note the following:

> The child's acquisition of formats and interactional routines is … a matter of her gradual socialization into the normative techniques of cultural life. Thus, recent research within the social interactionist perspective leads us to see how the child's linguistic development is inseparable from her socioaffective development; and this realization provides the study of linguistic development with one of the means to liberate it from the epistemological model which underpins generativism, and, equally, the classical view of scaffolding theory which holds that 'it is the interaction between LAD and LASS that makes it possible for the infant to enter the linguistic community' (Bruner, 1983a: 19; cf. Savage-Rumbaugh, Shanker and Taylor, 1998). For as Bruner emphasized in *Child's Talk*, and even more explicitly in *Acts of Meaning* (1990), the child's socialization into language demands its transformation into a cultural agent. (Shanker & Taylor, 2001, p. 59)

Note though this is an idea of viewing language as a form of social interaction with a normative character. Phipps and Gonzalez have a similar argument but one which is perhaps more relevant to the discussion in the context of learning 'another' language:

> Languaging is … inextricably interwoven with social experience – living in society – and it develops and changes constantly as that experience evolves and changes. The student of a language other than their own can be given an extraordinary opportunity to enter the languaging of others, to understand the complexity of the experience of others to enrich their own. (Phipps & Gonzalez, 2004, p. 3)

It is important to remember that Phipps and Gonzalez's view of languaging involves the use of multiple, independent languages, with 'the student of a language' having the 'extraordinary opportunity to enter the languaging of others'.

In the position with the relevant implications already expressed in the previous and present chapters of this thesis, there are no separate, different languages and there are no such constructs as learning English as a 'second', 'third' or 'foreign' language: there is only one human language. To paraphrase Albert Bandura (2008, p. 20), human language use, or languaging, is socially situated, highly contextualized and conditionally manifested. Adaptive language functioning requires both appropriate generalization in the face of bewildering situational variation and perceptive discrimination to avoid dysfunctional overgeneralization. Developing language users, therefore, vary in their languaging behaviour conditional on circumstances that reflect the diverse aspects of their lives. They engage, in other words, in languaculturing practices to wrestle with conflicting goals and courses of action, but these are instances of the same language being used to do different things under different life conditions, not different languages doing separate things.

The central argument is that the developing language user as a languaculturing (appreciated here intransitively) being is a dynamic meaning maker who is flexibly drawing upon a repertoire of linguistic resources to meet and make sense of the different situational demands and aspects of daily social life.

10.2.3.3 From complexity to supercomplexity

A recent conceptualization of language in Applied Linguistics is that it is a complex, dynamic system (see, e.g., the recent special issues of *Applied Linguistics*, 2006; *The Modern Language Journal,* 2008; and *Language Learning*, 2009). That is, of course, based on a view that there are different, independent languages in the world and that language learning entails a process of acquiring more than one complex linguistic system. When it is considered that only one language is used and shared among the entire human population in diverse sociocultural contexts, however, languaculturing becomes a supercomplex process, involving both receptive and productive use of sounds, forms and sub-systems as part of the meaning-making process in this age of supercomplexity (Barnett, 2000) and superdiversity (Blommaert & Rampton, 2011).

10.2.3.4 Simplicity

While languaculturing is supercomplex in this view, the key underlying principles of languaging, as suggested in this thesis, are fundamentally simple. Languaging involves a constant two-directional process of construction and reduction, with existing, acquired

structures, expressions and patterns and their corresponding meanings being continuously challenged (i.e., retained, modified or abandoned) and new ones formed so that at any point in time, the linguistic resources the languaculturing being enjoys are in a state of flux.

To summarize the discussion, the major shifts between conceptual frameworks from language learning to languaculturing associated with the position of this thesis may be represented as follows:

Table 10.1: From language learning to languaculturing

	Language learning	**Languaculturing**
Goal	Applying measurable linguistic knowledge	Making sense of the links between personal experiences and broader sociocultural and sociopolitical practices in, through and with languaging
Agent	Learner	The dynamic meaning maker as languaculturing being
Context	Classroom focus	Whole social world
Culture	Learning about	Living in and with
Performance	Prescribed by form	Freed through form
Position	Language at distance	Language from within
Concern	Overcoming deviant language use	Responding to and transforming dominant sociocultural and sociopolitical practices
Process	Error construction and eradication	Languaging practices with underlying construction and reduction principles
Outcome	Assessed performance based on proficiency in the target language	Human interaction and connection in, through and with languaging
Task	Complex	Supercomplex
Epistemology	Engagement *of* self and other	Critical reflective engagement *with* self and other

(Adapted from Phipps & Gonzalez, 2004)

As described in Table 10.1 (see also Section 10.4 for more discussion), engaging in languaculturing practices is a central part of what it means to be human. As Wittgenstein (1953, p. 23) famously put it, 'to learn a language is to enter a form of life'. Likewise, Phipps and Gonzalez reminded us a decade ago that

> *[t]o enter other cultures is to re-enter one's own*, understand the better the supercomplex variety of human experience … and become more deeply human as a result. (Phipps & Gonzalez, 2004, p. 3, italics original)

While these views are expressed within a multi-lingual frame of mind, the essence of the argument with a humanistic orientation and with a respect for the developing language user is not only shared in this thesis but is, significantly, a key theme running through the thesis.

Before we turn to the next research question, it may also be necessary to point out that some of the issues raised in relation to the position adopted in the thesis are open to debate. Viewing language as social interaction, for example, suggests the need to give more attention to the reception of texts (readers, intended audiences, etc.) than to the production of texts (see, e.g., McCarthy, 2001). The use of the term *developing language* might also raise questions. In the context of more conventional ways of thinking about 'learner language', the term might point to the possibility of a target end state of language learning, which is clearly incompatible with the position adopted in the thesis. Finally, the notion of complexity, as shown in this thesis, deserves more attention in its conceptualization in future research.

10.3 Research Question 4

> What does the study of a longitudinal corpus add to the literature of learner corpus research and to our understanding of human language development, including writing development, over time?

In this section, I attempt to answer the final research question by considering the findings from the five empirical studies presented in Chapters 4–8. I first offer two general observations, before I present more specific details of some main results from the analysis of the longitudinal corpus used in this thesis. (Note: I have changed the research question from the originally formulated 'second language development' and 'second language writing development' to the current 'human language development' and 'writing development', reflecting my position of rejecting a multi-lingual framing of language learning and use.)

10.3.1 General observations

1. The study of a longitudinal corpus has opened the possibility of addressing important unresolved questions in cross-sectional research

This seems obvious, but studies with a cross-sectional design have long been a main stream of research witnessed in the literature of both learner corpus research and SLA. An important question that has been seriously insufficiently addressed in learner corpus research is, 'How does the developing language change over time?'. On the other hand, a question which clearly requires much more attention from SLA researchers is, 'How do lexical-grammatical resources develop over time by a specific group of students?' (see though a few exceptions in the collection of papers in Schmitt (2004) and Barfield & Gyllstad (2009) on the acquisition of formulaic sequences and collocations).

While questions of the former kind suggest the need for longitudinal corpora, as 'they constitute one of the major desiderata for future learner corpus research' (Granger, 2012, p. 2), questions of the latter kind indicate how a lexical orientation towards language, fashioned by and most notably represented in the empirical, groundbreaking work of John Sinclair (e.g., 1991) and his associates (e.g., Francis et al., 1996, 1998), remains a Cinderella in SLA research with its main focus on grammatical development (see Chapter 2 for the relevant review). This latter case takes place despite the fact that vocabulary studies are flourishing in recent years (see, e.g., recent volumes such as Schmitt (2010) and Nation & Webb (2011) designed as guides for researchers and McCarthy et al. (2010) as a resource for language teachers).

Questions such as those presented above have been explored using the longitudinal data examined in this thesis, with the relevant details to be considered shortly.

2. The study of a longitudinal corpus can be expected to change the way of how the developing language may be perceived, with direct implications for research and pedagogical practice

As will be discussed further below, novel orientations such as the observation of two principles governing the language developmental process and the paradox of complexity evidenced in the developing language are dependent on the availability of longitudinal corpora. These observations in turn raise questions about the notion of complexity in the

study of the conventionally framed 'learner language' and about the very nature and process of language development itself.

All this has immediate implications for both research and the classroom. For the researcher, the changing nature of the developing language, as observed over time in the studies presented in this thesis, calls for a re-evaluation of the plausibility and scientific merit of research based on a methodology comparing 'learner language' with 'native-speaker language' (or with any other source of 'learner language', for that matter). Implied in the use of methodology of this kind, apart from a native-centric perspective, is the idea that the developing language is static, and so is the 'other' language ('native' or otherwise) being compared (and thus they are 'comparable'). This is certainly far from being true. The developing language, as shown in this thesis, is both dynamic and creative.

An account of the relevant pedagogical implications will be presented in Section 10.4.

10.3.2 Specific observations

What follows is a list of observations based on the study of the longitudinal corpus considered in this thesis.

1. Lexical and grammatical properties in the developing language grow and construct interaction with each other, resulting in regularities of association that form critically dynamic building ingredients for language development

Observed changes that take place in language development involves the interaction between grammatical and lexical resources. The findings emerging from the research support a theory of language on the interdependency of lexis and grammar (e.g., Sinclair, 1991, 2004; Francis et al., 1996, 1998; Carter & McCarthy, 1999; Hunston, 2003; Stefanowitsch & Gries, 2003), and are in line with usage-based theories of language acquisition (e.g., Tomasello, 2003; Bybee, 2008; N. Ellis & Cadierno, 2009; N. Ellis et al., 2013).

2. Complexity in the developing language is dynamic rather than static, thereby challenging the notion that language development is a process towards greater complexity in language use

The paradox of complexity has been observed in the developing language. In some way, the language is becoming more complex; in some way, it is becoming less complex. That is, what

is constructed in the developing language is complexity and what is reduced is also complexity. Language development then does not represent a natural progression from simple to more complex language use; languaging is intrinsically complex throughout the course of development, with different points in the developing language signifying complexity in different ways.

3. There is no single dominant pattern of development that is observable over time, but a complex series of them

Language development has been found to be an organic process whereby different language features exhibit different developmental patterns. The process may be characterized by four key developmental patterns over time: a three-phase pattern, an expansion pattern, a reduction pattern and no consistent pattern. This further supports the observation made above that language development cannot be expected to be a simple process of one's reaching or achieving greater complexity over time; it involves a range of language features with diverse growth patterns.

4. While all the observations paint a complex picture of how languaging dynamically unfolds over time, two fundamental processes or principles are at work in language development: *the construction principle* and *the reduction principle*

Language development is not just about accumulating and storing an ever larger pool of linguistic resources over time; as a living process, it also involves embracing reduction of certain linguistic forms, expressions and patterns of use. That is, both the construction and reduction principles are constantly at work in the course of language development, modelling the patterns of development at different points in time and simultaneously accounting for the complex series of developmental patterns underlying languaging practices.

5. Language development in instructional settings appears to be a process whereby differences in developmental patterns among individual students become significantly more pronounced over time

The developing language users are each different languaculturing beings. What is observed in this longitudinal investigation is that the difference in languaging practices between and among them becomes more distinct over time. While they begin not as a completely homogenous group of languaging beings, their productive language use has become much

more differentiated as time goes by. The greater variation, to a considerable extent, suggests the agentive nature of individual developing language users engaging in languaging practices.

6. Developing language users are versatile storytellers who are at times conforming to conventional, discursive constructions and at times displaying creative orientations in their ever changing narrative practices, just like any other human narrator

There is a critical engagement with distinct and discursive practices in the narrative discourse construction witnessed over time among the developing language users. Common and unique features have, for example, been observed across texts written at different points in time by the same developing language user. At the same time, both general and specific narrative practices are evidenced across texts written by different developing language users. These are clear signs of individuals' negotiation of meanings and relationships in the evolving personal and social worlds, and suggest the dynamic languaculturing practices in which they engage.

7. Language development, including writing development, is essentially a dynamic meaning-making process

The ever-changing nature of languaging practices observed over time signifies how the developing language user actively engages in the dynamics of meaning making. This involves not only the use of linguistic resources comprising lexis and grammar but also in the changing practices of narrative-discoursal construction. In other words, the developing language user is in a constant process of constructing and reconstructing meanings in, through and with languaging, and simultaneously changing the interplay between forms, meanings and languaging practices.

In the next section, we consider the practical consequences to the language teaching profession of an emphasis on, and a commitment towards, treating the developing language in its own terms.

10.4 Pedagogical considerations: From a deficit model to a languaculturing model

Historically, approaches to language learning and teaching have been based on a deficit model. In her discussion of the deficit theory used to interpret the linguistic characteristics of black and low socioeconomic status children three decades ago, Gail E. Thomas (1983, p. 108), for example, suggests how interpretations are made based on the assumption that 'there is a lack of stimulation of cognitive process (especially verbal) in poor children'. Citing deficit theorist

Jensen (1969) and Bernstein (1960), Thomas further explains how the perceived deficiencies in language used by these young people have been suggested to be accounted for by specific genetic factors (Jensen) or by factors relating to a specific social structural context (Bernstein).

New times present new challenges. If, as suggested above, there is now a growing recognition among scholars in the field that we need a re-conceptualization of the foundational assumptions (see, e.g., Klein, 1998; Cook, 2008; Ortega, 2013), so too, I would argue, it is time for a fresh look at how adequate the traditional tools of the trade are for addressing the new questions now emerging and facing the language teaching profession. This section proposes an alternative model introduced earlier in Section 10.2, the languaculturing model, which aims to revitalize how practitioners and policymakers go about thinking and acting to promote an approach that redresses the balance between the deficit and proposed models.

As noted before, the term *languaculturing* is based on the conception of 'languaculture' introduced by the anthropologist Michael Agar, and the languaculturing model proposed here draws upon theoretical insights, empirical work and recommendations from a number of disciplines, spanning medical sociology (the asset model), management theory (Appreciative Inquiry) and applied linguistics. Unlike the deficit model which tends to define the languaging being in negative terms (e.g., as a 'defective communicator', see Chapter 1), the languaculturing model focuses on highlighting positive capability as a way to promote a critical awareness of the linguistic-cultural resources the dynamic meaning maker is developing. The discussion here draws heavily from Morgan and Ziglio (2007), Wells (2009), Bushe (2013) and García (2014).

10.4.1 Principles

The languaculturing model is based on the assumptions (1) that language teaching and learning as an area or process of enquiry is a socially constructed phenomenon, which has no tangible reality, and (2) that ways of appreciating and managing this process are limited only by human imagination and the agreements people make with each other. The model is based on the following principles, first introduced by Cooperrider and Whitney (2005) in their work on Appreciative Inquiry:

1. **The constructionist principle** proposes that what we believe to be true determines what we do, and thought and action emerge out of relationships. Through languaging and discourse

of day to day interactions, people co-construct the learning and teaching process. The purpose of learning and teaching in the classroom is to stimulate new ideas, stories and images that generate new possibilities for action.

2. **The principle of simultaneity** proposes that as we enquire into human communication systems, we change them. The seeds of change, the things people think and talk about, what they discover and learn, are implicit in the very first questions asked. Questions are never neutral, they are fateful, and social systems move in the direction of the questions they most persistently and passionately discuss.

3. **The poetic principle** proposes that the classroom life is expressed in the stories people tell each other every day, and the story of the classroom is constantly being co-authored. The words and topics chosen for enquiry and discussion have an impact far beyond just the words themselves. They invoke sentiments, understandings and worlds of meaning. In all phases of the enquiry, effort is put into using words that point to, enliven and inspire the best in people.

4. **The anticipatory principle** posits that what we do today is guided by our image of the future. Human systems are forever projecting ahead of themselves a horizon of expectation that brings the future powerfully into the present as a mobilizing agent. The languaculturing model uses artful creation of positive imagery on a collective basis to refashion anticipatory reality.

5. **The positive principle** proposes that momentum and sustainable change requires positive affect and social bonding. Sentiments like hope, excitement, inspiration, camaraderie and joy increase creativity, openness to new ideas and people, and cognitive flexibility. They also promote the strong connections and relationships between people, particularly between groups with different power relations, required for collective enquiry and change.

10.4.2 Developing a languaculturing model

The following is a list of characteristics associated with the model, with a consideration of how one may go about activating a languaculturing approach to language teaching and learning.

1. At the individual level: individual students are critically aware that any deviant use of language is not an error, but part of their own meaning-making resources, which are continually shaped by their unique personal biographies in interaction with the immediate, sociocultural contexts. There is a constant reflective engagement with the idea that a 'language error' comes from the dominant sociocultural and sociopolitical practices; scientifically it is untenable. When deviant language use is produced or observed, that is a sign of languaging practices away from the dominant language practice (intentionally or otherwise), and that is all about it: it is not a sign of one's being 'dull' or 'slow' in 'acquiring' the language as is often perceived in many classrooms today. In other words, positive values and self-esteem of the students are emphasized. In a classroom based on the languaculturing model, students feel good as human participants in the learning and teaching process, are aware of their languaging practices, and feel useful and morally valuable.

The teacher, responding and enacting within a languaculturing model, engages in the following:

(a) to treat what the student has to say and write as worthy of careful attention;
(b) to do one's best to understand what he or she means and offer (but not enforce the use of) alternatives of expressions commonly employed in the dominant languaging world;
(c) to take the student's meaning as the basis for what one says next;
(d) in selecting and encoding one's message, to take account of the student's comprehension of the message. Here a communication breakdown signifies not necessarily the student's failure to comprehend the message. What is emphasized here is that the student is an equally respectable partner in the two-way communication process. There are times when the student fails to understand the teacher and there are times when the teacher fails to understand the student, which is characteristic of any other ordinary communication that takes place in real life. What is important is negotiation of meanings between each other and a deep and equal respect for each other, including the individual expressions.

2. At the institutional and/or policymaking level: languaging practices among diverse student groups are valued and appreciated. This is no doubt challenging. Writing in relation to the education of minority students in the US, Cummins (1986/2001, p. 673) observes the following, which is equally relevant and vital to our discussion today as it was 30 years ago:

> … educators and policymakers are faced with both a personal and a political challenge. Personally, they must redefine their roles within the classroom, the community, and the broader society so that these role definitions result in interactions that empower rather than disable students. Politically, they must attempt to persuade colleagues and decisionmakers — such as school boards and the public that elects them — of the importance of redefining institutional goals so that the schools transform society by empowering … students rather than reflect society by disabling them.

Underlying the argument here is a social justice issue.

10.4.2 .1 The importance of languaculturing

Challenges aside, the languaculturing model adds value to the deficit model by:

1. holding the promise of promoting the languaging practices of developing language users that would enable them to value themselves in schools, while supporting a social justice agenda that holds developing language users as knowers, thinkers and dynamic meaning makers;
2. returning the agency and responsibility for learning to the developing language users themselves, enabling them to control their school languaging practices in ways that allow them to think more deeply and to engage and invest in their learning, and freeing them from being constrained by a target-language-only policy;
3. promoting the developing language user population as a co-producer and co-participant of languaging rather than simply a passive recipient of linguistic knowledge, thereby reducing the demand on the teacher as a 'knower' or 'bearer' of linguistic knowledge;
4. strengthening the capacity of individuals and communities to realize their potential for contributing to and developing a mutual respect for diverse languaging practices; and
5. contributing to more equitable and sustainable educational and sociocultural development.

In reality, both deficit and languaculturing models may be important. A deficit model might be more relevant, or even desired, in a context requiring language needs analysis or with expectations for an approximation to a targeted 'norm' of language use, as pursued or expressed by some students for various personal and social reasons (see, e.g., Timmis, 2002). More work, however, needs to be done to redress the balance between the more dominant deficit model and the much less well-known (and understood) languaculturing model. The proposed model presented here promotes a more humanistic approach to understanding the

theory and practice of languaculturing as a way to enhance social and educational equity. In doing so, it has the potential to create a more robust evidence base that demonstrates why investing in the languaculturing practices of individuals, communities and institutions can help to reduce the language gap between those traditionally perceived as linguistically and culturally disadvantaged in society and those who achieve the best social (and economic) capital.

At its core, the languaculturing model asks:

- What external factors contribute to an appreciation of diverse languaging practices?
- What factors help developing language users to be more able to critically reflect on and engage with personal and broader, socially dominant languaging practices?
- What opens us to more fully experience (human) life?
- What produces overall levels of positive feelings and self-esteem among developing language users who are engaging in languaculturing practices?

Applying this model to the search for evidence on the opportunities and constraints in languaging practices and the evidence of the most effective actions has the potential to explain further what is required to address social and educational inequities. It also encourages the discipline to move towards finding answers to what enriches the developing language as a meaning-making resource, rather than its traditional focus of generating evidence about the causes and distribution of 'learner errors' and about the questionable notion of 'fossilization'. The languaculturing model therefore calls for a rethinking of the theoretical basis on which language teaching and learning is built.

10.5 Time for a scientific revolution?

What is proposed here is, to some extent, a utopian vision. However, it is, I hope, a practical utopia. Just as how Appreciative Inquiry has had a profound impact on organization development practice around the world in business, non-profit, and governmental organizations as well as communities (Bushe, 2013), and how the asset model has stimulated a large and still growing number of researchers, policymakers and practitioners to think differently about how they approach their goal of improving the health of populations away from a deficit model (Morgan & Ziglio, 2007), the proposed model discussed here suggests that it is possible to step outside the well-worn frameworks of thought and practice.

Adopting a languaculturing approach fundamentally challenges the arbitrarily, politically and historically defined boundaries between 'languages', and between 'language' and 'culture'. In this view, one is no longer 'learning' language in the traditional sense: one is living in and with languaculturing practices to make sense of one's personal and social world.

Languaculturing as a model for conceptualizing and activating languaging practices is part of, as Phipps and Gonzalez (2004, p. 166) notes in their discussion of languaging and being intercultural, 'a political project of engagement with questions of power and of being human'. In fact, their observations (Phipps & Gonzalez, 2004, p. 167) are so relevant to the present argument that it is worth quoting what they say in the next two paragraphs, although admittedly the observations are re-contextualized here for the purpose of our present discussion.

Languaculturing is a way of being, encompassing the whole social world. It is at the same time transformational. It is in and of itself embodied knowledge. Through languaculturing, people come to make sense of, and to shape, their worlds. They become active agents in creating their human and material environments. Languaculturing is inextricably relational. It is a social way of being. As Spivak (1999, p. 27) reminds us, 'No speech is speech unless heard'.

As languaculturers, we are people who move in and through words as actions, who develop and change constantly as the experience of languaculturing evolves and changes us. A languaculturing student and a languaculturing teacher are given a unique opportunity to enter the languaculturing of others, to open up the ways in which the complexity and experience of others may enrich life and in which a greater understanding and engagement with self and other may be achieved.

It should perhaps be noted at this point that the discussion thus far on the need to give a deep and equal respect to the developing language user and to the associated languaging practices is presented primarily through a sociocultural lens. Remarkably recent developments in cognitive science point to similar insights in the context of thinking about children, childhood, development, and the role that the so-called 'immaturity' has played in making us the species, and the people, that we are.

In a highly acclaimed text, *Why Youth is* Not *Wasted on the Young*, for example, David F. Bjorklund (2007) shows that contrary to the common assumption that the purpose of childhood is to set the stage for adulthood, this phase in human life has an integrity of its own. In Bjorklund's words,

> Children have ways of learning and knowing, through discovery and play, that permit them to master new skills easily and to invent new knowledge (such as language if the conditions are right). We owe a substantial debt to our childhoods. We could not be the adults we become if children were more like us in how they see and learn about the world. … This view of development makes childhood more than a training school for adulthood. Childhood has an integrity of its own. Children are not simply little, incomplete adults, but people whose minds and behaviors are well suited to the demands of their young lives. It should be our jobs as adults to promote children's development, putting their youthful ways of learning and knowing to good use, not only to prepare them for the adults they will become, but *to foster the children they are*. (Bjorklund, 2007, p. 223, emphasis added)

As has been emphasized since Chapter 1, developing language users are not deficient native speakers. Their languaging practices need to be fostered rather than to be shaped according to an ideal native-speaker norm. To the extent that it is able to develop such a capacity on a significant scale, Applied Linguistics as a profession is likely to move from being either the blunt tool of policymakers on the one hand or the arcane scholarly endeavour of academics on the other, to being a powerful force for change in the way we think about language, education and the personal and social worlds of young people. I believe that a view which prioritizes the dynamic and supercomplex linguistic-cultural resources of the developing language user can make a meaningful contribution to the urgently needed re-examination of the goals of Applied Linguistics in the contemporary globalized world: from aiming to address language-related, real-life problems to transforming these problems into significant questions that can be publicly addressed. Appropriately, if *anti*-grammatically, termed 'languaculturing', this new orientation might be the first step in a scientific revolution.

10.6 Conclusion

This chapter has considered the extent to which it is possible to research and analyze the developing language in its own terms when studying the language developmental process. In doing so, two perspectives are discussed, together with two metaviews presented. While it

was noted that the position I am associated with is just a perspective, it is important to note that it is also the perspective that I choose. The contributions of the study of a longitudinal corpus to learner corpus research and to our understanding of human language development have also been considered.

I have argued that we need to go beyond a deficit model in research and pedagogical practice. While practices based on a deficit model might be important and necessary to identify levels of needs and priorities in certain contexts, they do not go far enough in allowing developing language users to press their meaning-making resources, afforded by their distinct linguistic and cultural backgrounds, into valued, communicative use. These practices need to be complemented by an alternative model, one which I have termed *languaculturing*.

This chapter has suggested that a more professionally rewarding pedagogical practice might take place in a classroom where:

- individual students are critically aware that signs of languaging practices away from the dominant language practice (intentionally or otherwise) are not signs of their being 'dull' or 'slow' in 'acquiring' the language;
- students feel good as human participants in the learning and teaching process and are aware of their languaging practices, and feel useful and morally valuable;
- the teacher has an equal respect for what the student has to say and write;
- the teacher offers, but does not enforce the use of, alternatives of expressions commonly employed in the dominant languaging world;
- the teacher takes the student's meaning as the basis for what he or she says next; and
- the teacher engages in an equal negotiation of meanings with the student and does his or her best to understand what the student intends to mean, and so does the student.

In short, what is emphasized here is a two-directional process of human growth through which mutual languaculturing enrichment as well as a critical engagement with social reality is made possible (cf. Phipps & Gonzalez, 2004).

A languaculturing model, informed by insights from a number of disciplines, has the potential to encourage the field to move forward by helping policymakers, researchers and practitioners to rethink how to conceptualize languaging practices in order to:

- raise the self-esteem and positive values of individual students as dynamic meaning makers;
- ensure that all policies and programmes aimed at tackling educational inequities take account of the valuable linguistic-cultural resources already existing in individuals and communities; and
- sharpen the efficiency of schools and institutions to contribute to the overall well-being of the communities they serve.

In making a case for a languaculturing approach, I hope that I have not given the impression that the deficit model deserves little attention. As noted on at least two occasions, the model might be useful in specific contexts where there are some students who aspire to a 'native-speaker' model of language use. What I am arguing, however, is that there is an urgent need to redress the balance between the proposed languaculturing model and the more dominant deficit model for research and practice. This could help us to unlock some of the existing barriers to effective action relating to applied linguistic concerns, and to better understand the factors that influence languaging practices and what can be done about them. All this would promote a positive and inclusive approach to action.

Chapter 11

Conclusion

I am in competition with no one. I have no desired to play the game of being better than anyone. I am simply trying to be better than the person I was yesterday. (Languaculturer, 2013)

11.1 Introduction: Does one study apples according to the norm of oranges?

I am simply trying to be better than the person I was yesterday. Such are the powerful words of a languaculturer, who so happened to be one of the students in 2007 contributing their texts to the making of my current longitudinal corpus. The words were quoted directly from his Facebook post in 2013 and are shown here, with his permission. They represent a voice which reveals a great deal about one's positioning of oneself, about one's making sense of one's life, about one's shifting sense of self-renewal, and about human agency in general. Any reader of this thesis will no doubt notice, too, how I have changed my own perspectives over time, from the beginning of this research to its completion, engaging in 'self renewal' as a researcher trying to make sense of human language use and language development,

The message here is important for researchers and practitioners who are working with students. With their best intentions, the researchers might reconsider ways of approaching their studies involving language users (e.g., Do we compare one individual or group of students with another individual or another group? But how do we account for those students who, like the languaculturer noted above, aim to be '*I am in competition with no one. … I am simply trying to be better than the person I was yesterday*'?). On the other hand, in the best interests of their students, the practitioners might reflect on ways of engaging the students in their classroom, especially those who have no desire 'to play the game' of being better than others. Some of these ways have been suggested in the previous chapters for both researchers and practitioners and will be briefly revisited in this concluding chapter. I must also note that the distinction made here between a researcher and a practitioner is certainly not always true.

This final chapter first presents a summary of the findings based on LoCDeLUNT, or what I now prefer to call the Longitudinal Corpus of Languaculturer Narrative Texts (LoCLaNT) and of the pedagogical implications arising from the understanding that the language user is a

dynamic meaning maker: a languaculturer. It then considers some contributions made by this thesis, from theoretical, empirical, methodological and practical perspectives, before pointing to some productive avenues for further research.

11.2 Summary of research findings

Using the tools and methods from corpus linguistics and written discourse analysis, the research reported in this thesis has explored the notion of change in the developing language. A number of important findings have been observed in the study of LoCLaNT. They are presented as follows:

1. Lexical and grammatical properties in the developing language grow and construct interaction with each other, resulting in regularities of association that form critically dynamic building ingredients for language development.
2. Complexity in the developing language is dynamic rather than static, thereby challenging the notion that language development is a process towards greater complexity in language use.
3. There is no single dominant pattern of development that is observable over time, but a complex series of them.
4. While all the observations paint a complex picture of how languaging dynamically unfolds over time, two fundamental processes or principles are at work in language development: *the construction principle* and *the reduction principle*.
5. Language development in instructional settings appears to be a process whereby differences in developmental patterns among individual students become more pronounced over time.
6. Developing language users are versatile storytellers who are at times conforming to conventional, discursive constructions and at times displaying creative orientations in their ever changing narrative practices, just like any other human narrator.
7. Language development, including writing development, is essentially a dynamic meaning-making process.

11.3 Summary of pedagogical implications

1. The goal of language teaching is, based on a languaculturing model, to help students to make sense of the links between their personal experiences as shaped by individual life

histories, and the broader sociocultural and sociopolitical practices in, through and with languaging.

2. Students are not passive recipients of linguistic knowledge but dynamic meaning makers, or in one word, languaculturers.
3. The classroom focus is not only on the classroom life but also on the whole social world.
4. Students are supported not just to learn about others' cultural practices but to live in and with such practices.
5. Students' languaging practices are valued not based exclusively on the conventionally determined forms but on innovative or creative expressions at the same time.
6. There is a critical awareness among students that languaging practices are not about learning an 'external' language, but rather about valuing what one has been using and continually changing, developing and renewing.
7. The concern of the language teaching and learning process is not about addressing deviant language use but rather about responding to and transforming dominant sociocultural and sociopolitical practices.
8. There is a critical awareness among students and teachers that daily languaging practices are not about noticing 'errors' and addressing them, but rather about appreciating languaging practices with underlying construction and reduction principles.
9. The outcome of the language teaching and learning process is not about assessed performance based on proficiency in the target language but about human interaction and connection in, through and with languaging.
10. There is a critical awareness among students and teachers that languaculturing is a supercomplex process involving and requiring an equally deep appreciation for such a practice and process from schools and communities.
11. In achieving all this, a constant critical reflective engagement *with* self and other is essential.

All this suggests a huge challenge ahead of the pedagogical applied linguist; perhaps it is important to remember the following familiar lines:

> Politically, they must attempt to persuade colleagues and decisionmakers — such as school boards and the public that elects them — of the importance of redefining institutional goals so that *the schools transform society by empowering … students rather than reflect society by disabling them*. (Cummins, 1986/2001, p. 673, emphasis added)

The two assumptions on which the languaculturing model is based might also be rehearsed: (1) that language teaching and learning as an area or process of enquiry is a socially constructed phenomenon, which has no tangible reality, and (2) that ways of appreciating and managing this process are limited only by human imagination and the agreements people make with each other.

11.4 Contributions of this thesis

Overall, the longitudinal investigation into the languaging practices of the developing language users considered in this thesis contributes in a number of important ways to learner corpus research, SLA and the broader field of Applied Linguistics.

On a theoretical level, this thesis contributes to the current debate on views of 'learner' and 'learner language', and challenges a construct lying at the heart of much applied linguistic research: the notion of accuracy. I have argued that 'accuracy' is a socially constructed concept: it has no real scientific basis. In approaching the task of researching language development, this thesis has considered all the relevant instances of use, conventional as well as innovative, of the developing language as meaningful expressions by the developing language users, and given these instances equal consideration in the analysis. The outcome, with the removal of this cultural layer of preconception, shows that the developing language is a much more exciting site for linguistic exploration than might otherwise be the case through a mere prescriptive analysis of 'accurate' and 'inaccurate' language use.

All this is part of an 'all-instances-are-meaningful' orientation towards the developing language. Significantly, this perspective contributes to new observations such as the underlying principles of languaging. It also has the practical value of stimulating new thinking about relevant issues of usefulness or adequacy about current constructs relating to conceptualizations of 'learners' and 'learner language' which, in turn, suggest theoretical refinements. The proposed notions as *dynamic meaning makers, languaculturers, the developing language, languaging* and *languaculturing* might serve the Applied Linguistics community better in the foreseeable future.

On an empirical level, because learner corpus research is so far largely cross-sectional in design and SLA is in continuous search for sizeable data sets to make generalizations of

findings possible, the results from the series of studies reported in this thesis, based on a longitudinal corpus involving a group of 124 developing language users, add to the current research base in both fields. The thesis contributes to the empirical record of ways in which language resources involving patterns of use of the selected lexis change over time (Chapters 4–6), and of whether and to what extent individual users' narrative texts expand or contract in length (Chapter 7), or change in structure (Chapter 8), as time goes on. It contributes, at the same time, to the empirical record of changing patterns of language use and development in both group performance (Chapters 4–6) and individual performance (Chapters 7–8) among the under-researched young, secondary school students considered in this thesis.

Most importantly, the thesis is itself an empirical record, within the space available, of instances exemplifying the developing language users as dynamic meaning makers through their deployment of changing resources comprising lexis and grammar (Chapters 4–6) and texts (Chapter 8) as meaningful units of language use. All this also suggests the agentive nature of human languaging practices and development.

On a methodological level, this thesis makes a direct contribution to three research communities: learner corpus research, SLA and second language writing. To the learner corpus research community, the thesis exemplifies an approach to corpus investigation that prioritizes the study of internal changes in, rather than involves external comparisons of, the research data, leading to the discovery of new insights into the working of the developing language (Chapters 4–6). To the SLA community, this thesis uses individual words, rather than grammatical structures, to track language development, thereby contributing to an inductive, corpus approach to the study of language development (Chapters 4–6). To the second language writing research community, it adopts a matrix analysis to investigate the changing structures of the individual narrative texts over time, contributing a novel methodology in the study of writing development that is based on the innovative application of a discourse analytical approach (Chapter 8).

All this essentially points to the central methodological and epistemological contribution of this thesis: the study of the developing language in the process of change, through which (1) the essence of language development can be captured and the nature of the developing language revealed, and through which (2) the developing language is given a deep and equal

respect that is long overdue, no longer just an imperfect version of the 'native-speaker' language. One does not study apples according to the norm of oranges.

On a practical level, the findings of the research reported in this thesis on the nature of the developing language and the process of language development can be disseminated throughout the broader applied linguistic community, including language teaching practitioners, materials writers, syllabus designers, and curriculum developers and policymakers, as well as to the general public interested in issues and debates in language learning and teaching. The real contribution of this thesis is, however, not so much in mere sharing of the findings but rather, in inviting a change, through the findings, in the ways we think about, understand and respond to some long-held, socially constructed and ideologically based assumptions about language use. The issue of accuracy is a notable case in point.

On the one hand, the concern is real, even within our professional and research community, that some instances of use of 'learner language' are considered 'accurate' while others are not; it is also true, on the other hand, that such a view is based on social preconceptions which have no real scientific basis. As evidenced in the research reported (in, e.g., Chapter 5), once these ideological preconceptions are set aside and all instances of use of the developing language are observed as they are, we begin to see and appreciate fundamental principles governing the nature of languaging practices. In other words, changing the way we perceive things from a prescriptive perspective (imperfect language use by 'learners') to a descriptive perspective (natural language use by meaning makers) has an important, practical outcome of advancing our understanding of human language development. It should be noted though that both views are ideologically based. With this knowledge in mind about the relative merits of each perspective, however, more informed decisions can be made at the levels of pedagogy, materials production, syllabus design, curriculum planning and public debate.

11.5 Implications for further research

This thesis makes use of corpus evidence of written output to study language development. Inevitably issues such as the nature of input, intake and interaction, and the role and manner of feedback that the students have received are left unaddressed. An informant-based study might also have enriched our view of the data: for example, how many language users might agree or disagree with the languaculturer quoted at the beginning of this chapter who saw himself as mainly concerned with self-development and 'in competition with no one' (see,

e.g., Timmis, 2002)? Questions such as L1 transfer might also be raised. When framed within a languaculturing perspective, however, language forms, it must be pointed out, become less of a concern and the dynamic meaning maker can be observed to be flexibly drawing upon a repertoire of linguistic resources to meet and make sense of the different situational demands and aspects of daily social life.

The research and the research methodology reported in this thesis both open up considerable scope for further work. First of all, using function words as the starting point for linguistic investigation holds much promise for productive research enquiry, as exemplified in the series of three corpus studies considered here. Future research is likely to benefit a great deal from a lexical approach of this kind. Second, the matrix analysis yields interesting findings about the changing narrative practices of the developing language users. This is a discourse analytical tool that deserves attention from those working in the fields of second language writing research and language teaching research.

Next, and perhaps rather obvious, more longitudinal research projects need to be developed so that more meaningful contribution to the field can be made. As shown in this thesis, the developing language can be demonstrably shown to be dynamic and living simply because the thesis draws upon empirical, longitudinal data, with different sub-sets of data points capturing and revealing different moments of language use. Interested researchers might refer to Ortega and Byrnes (2008) for useful points and advice on how to go about conducting longitudinal studies.

Finally, as this thesis indicates, how one approaches the data has direct impact on the outcome of one's research. The paradox of complexity is a compelling example, observable only when all the instances of use in the developing language are considered as equally valid features of natural human language. This brings us back to the point on the research cycle involving one's theoretical orientation and practice. To be equally respectful of the developing language, or not to be? That is the real question.

11.6 Conclusion

This thesis has considered a number of issues, from theory and method in Corpus Linguistics to research and pedagogical practice in Applied Linguistics. Many more issues that deserve investigation could be raised, such as the one-language hypothesis with its implications for

the study of semantic sequences (e.g., Hunston, 2006; Groom, 2007), which may be broadened in conceptualization now to include not only alternative expressions in ‘English’, but also a greater number of forms of expressions from ‘other languages’. A thesis must, however, stop somewhere.

In a fast-changing field like Applied Linguistics, what one foresees is a continual advancement in both theory and practice. The languaculturing model proposed in this thesis, I hope, is part of this wave of exciting change, drawing disciplines such as medical sociology, management theory and linguistic anthropology closer together, with the value of enriching our understanding of human beings in all of their expressions.

References

Abd. Samad, A., Hassan, F., Mukundan, J., Kamarudin, G., Syd. Abd. Rahman, S. Z., Md. Rashid, J., & Vethamani, M. E. (2002). *The English of Malaysian school students (EMAS) corpus*. Serdang: Universiti Putra Malaysia Press.

Agar, M. (1994). *Language shock: Understanding the culture of conversation*. New York, NY: Quill.

Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of *make* in native and non-native student writing. *Applied Linguistics, 22*(2), 173-194.

Anderson, P. W. (1972). More is different. *Science, 177*, 393-396.

Babcock, B. A. (1977). The story in the story: Metanarration in folk narrative. In R. Bauman (Ed), *Verbal art as performance* (pp. 61-80). Prospect Heights, IL: Waveland.

Bakhtin, M. M. (1981). *The dialogical imagination: Four essays by MM Bakhtin*. C. Emerson & M. Holquist (Trans.). M. Holquist (Ed.). Austin, TX: University of Texas Press.

Bandura, A. (2008). Toward an agentic theory of the self. *Advances in self research*, *3*, 15-49.

Bardovi-Harlig, K. (2000). *Tense and aspect in second language acquisition: Form, meaning and use*. Malden, MA: Blackwell.

Barfield, A., & Gyllstad, H. (Eds.) (2009). *Researching collocations in another language: Multiple interpretations*. Basingstoke: Palgrave Macmillan.

Barnett, R. (2000). *Realizing the university in an age of supercomplexity*. Philadelphia: Open University Press.

Barnett, R. (2013). Potentials and risks of collaboration: Two sides of the same coin or the same side of the coin? *Policy Futures in Education, 11*(4), 449-452.

Bauman, R., & Briggs, C. L. (1990). Poetics and performance as critical perspectives on language and social life. *Annual Review of Anthropology, 19*, 59-88.

Bailey, N., Madden, C., & Krashen, S. (1974). Is there a "natural sequence" in adult second language learning? *Language Learning, 24*(2), 235-243.

Behrens, H. (2008). Corpora in language acquisition research: History, methods, perspectives. In H. Behrens (Ed.), *Corpora in language acquisition research: History, methods, perspectives* (pp. xi-xxx). Amsterdam: Benjamins.

Bernstein, B. (1960). Language and social class. *British Journal of Sociology, 11*(3), 271-276.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.

Bitchener, J. (2012). A reflection on 'the language learning potential' of written CF. *Journal of Second Language Writing, 21*(4), 348-363.

Bjorklund, D. F. (2007). *Why youth is not wasted on the young: Immaturity in human development*. Malden, MA: Blackwell.

Bley-Vroman, R. (1983). The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning, 33*(1), 1-17.

Blommaert, J. & Rampton, B. (2011). Language and superdiversity. *Diversities 13*(2), 1-22.

Blum-Kulka, S. (2005). Modes of meaning making in young children's conversational narratives. *The sociolinguistics of narrative* (pp 149–171). Antwerpen: John Benjamins.

Boers, F., & Lindstromberg, S. (Eds.) (2008). *Cognitive linguistic approaches to teaching vocabulary and phraseology*. Berlin: Mouton de Gruyter

Bogaards, P., & Laufer, B. (Eds.) (2004). *Vocabulary in a second language: Selection, acquisition and testing*. Amsterdam: John Benjamins.

Bushe, G. (2013). The Appreciative Inquiry model. In E. Kessler (Ed.), *The encyclopedia of management theory* (pp. 42-45). Thousand Oaks, CA: SAGE.

Bybee, J. (2008). Usage-based grammar and second language acquisition. In P. Robinson & N. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 216-236). New York: Routledge.

Byrnes, H., Maxim, H., & Norris, J. M. (2010). Special issue: Realizing advanced foreign language writing development in collegiate education: Curricular design, pedagogy, assessment. *Modern Language Journal, 94*(s1).

Carter, R. (1998). *Vocabulary: Applied linguistic perspectives* (2nd edn). London: Routledge.

Carter, R. A., & McCarthy, M. J. (1999). The English get-passive in spoken discourse: Description and implications for an interpersonal grammar. *Journal of English Language and Linguistics, 3*(1), 41-58.

Carter, R., & McCarthy, M. (2006). *Cambridge grammar of English: A comprehensive guide to spoken and written grammar and usage*. Cambridge: Cambridge University Press.

Charles, M. (2006). The construction of stance in reporting clauses: A cross-disciplinary study of theses. *Applied Linguistics, 27*(3), 492-518.

Cheng, W., Greaves, C., & Warren, M. (2006). From n-gram to skipgram to concgram. *International journal of corpus linguistics*, *11*(4), 411-433.

Chernuisheva, I. I. (1964). *Die Phraseologie der gegenwärtigen deutschen Sprache*. Moscow: Vuisshaya Shkola.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Clahsen, H. (Ed.) (1996). *Generative perspectives on language acquisition.* Amsterdam: Benjamins.

Cobb, T. (2003). Analyzing late interlanguage with learner corpora: Quebec replications of three European studies. *The Canadian Modern Language Review, 59*(3), 393-423.

Collins, P. (2009). *Modals and quasi-modals in English*. Amsterdam: Rodopi.

Cook, V. (1992). Evidence for multicompetence. *Language learning, 42*(4), 557-591.

Cook, V. (2008). Multi-competence: Black hole or wormhole for second language acquisition research? In Z. Han (Ed.), *Understanding second language process* (pp. 16–26). Clevedon, UK: Multilingual Matters.

Cook, V. (2012). Multicompetence. In C. A. Chapelle (Ed.), *The encyclopedia of Applied Linguistics*. Oxford: Wiley-Blackwell.

Cooperrider, D.L., & Whitney, D. (2005). A positive revolution in change: Appreciative Inquiry. In D. L. Cooperrider, P. Sorenson, T. Yeager, & D. Whitney (Eds.), *Appreciative Inquiry: Foundations in positive organization development* (pp.9-33). Champaign, IL: Stipes.

Corder, S. P. (1967). The significance of learner's errors. *International Review of Applied Linguistics in Language Teaching, 5*(4), 161-170.

Corder, S. P. (1971). Idiosyncratic dialects and error analysis. *International Review of Applied Linguistics in Language Teaching, 9*(2), 147-160.

Corder, S. P. (1981). *Error analysis and interlanguage*. Oxford: Oxford University Press.

Cummins, J. (1986). Empowering minority students: A framework for intervention. *Harvard Educational Review, 56*(1), 18-36. Reprinted in Cummins, J. (2001).

Darwin, C. (1877). A biographical sketch of an infant. *Mind, 2*, 285-94.

Davies, A., & Elder, C. (2004). General introduction: Applied Linguistics: Subject to discipline? In A. Davies & C. Elder (Eds.), *Handbook of Applied Linguistics* (pp. 1-15). Oxford: Blackwell.

de Certeau, M. (1984). *The practice* of everyday *life*. Berkeley: University of CaliIornia Press.

De Cock, S. (1998). A recurrent word combination approach to the study of formulae in the speech of native and non-native speakers of English. *International Journal of Corpus Linguistics, 3*(1), 59–80.

De Cock, S., Granger, S., Leech, G., & McEnery, T. (1998). An automated approach to the phrasicon of EFL learners. In S. Granger (Ed.), *Learner English on computer* (pp. 67-79). London: Longman.

Deleuze, G. (1994). *Difference and repetition*, trans. P. Patton. New York, NY: Columbia University Press.

Doughty, C. (1991). Instruction does make a difference: The effect of instruction on the acquisition of relativization in English as a second language. *Studies in Second Language Acquisition, 13*(4), 431-469.

Dulay, H., & Burt, M. (1973). Should we teach children syntax? *Language Learning, 23*(2), 245-258.

Dulay, H., & Burt, M. (1974). Natural sequences in child second language acquisition. *Language Learning, 24*(1), 37-53.

Ellis, N. C. (2011). The emergence of language as a complex adaptive system. In J. Simpson (Ed.), *Handbook of Applied Linguistics* (pp. 666-679). London: Routledge

Ellis, N. C., & Cadierno, T. (Eds). (2009). Constructing a second language. *Annual Review of Cognitive Linguistics, Special section. 7*, 111-290.

Ellis, N. C., & Larsen-Freeman, D. (2009). Constructing a second language: Analyses and computational simulations of the emergence of linguistic constructions from usage. *Language Learning*, *59*(s1), 90-125.

Ellis, N. C., O'Donnell, M. B., & Römer, U. (2013). Usage-based language: Investigating the latent structures that underpin acquisition. *Language Learning, 63*(s1), 25-51.

Ellis, R. (1989). Are classroom and naturalistic acquisition the same?: A study of the classroom acquisition of German word order rules. *Studies in Second Language Acquisition, 11*(3), 305-328.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Ellis, R. (2008). *The study of second language acquisition* (2nd edn). Oxford: Oxford University Press.

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Ocford: Oxford University Press.

Firth, A., & Wagner, J. (1997). On discourse, communication, and (some) fundamental concepts in SLA research. *The Modern Language Journal*, *81*(3), 285-300.

Firth, J.R. (1957). *Papers in Linguistics* 1934-1951. London: Oxford University Press.

Firth, J.R. (1968). A synopsis of linguistic theory, 1930-55. In F. R. Palmer (Ed.), *Selected papers of J. R. Firth 1952-1959* (pp. 168-205). London: Longman.

Fitzpatrick, T., & Barfield, A. (Eds.) (2009). *Lexical processing in second language learners*. Bristol: Multilingual Matters.

Francis, G. (1993) A corpus-driven approach to grammar: Principles, methods and examples. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 137–156). Amsterdam: Benjamins.

Francis, G., Hunston, S., & Manning, E. (1996). *Collins Cobuild Grammar Patterns 1: Verbs*. London: HarperCollins.

Francis, G., Manning, E., & Hunston, S. (1998). *Collins Cobuild Grammar Patterns 2: Nouns and adjectives*. London: HarperCollins.

Francis, W. N. (1992). Language corpora BC. In J. Svartvik (Ed.), *Directions in corpus linguistics: Proceedings of Nobel symposium 82* (pp. 17–34). Berlin, Germany: De Gruyter.

Fries, C. C. (1952). *The structure of English*. New York: Harcourt, Brace & World.

Fromm, J. (2004). *The emergence of complexity*. Kassel: Kassel University Press.

García, O. (2014). TESOL translanguaged in NYS: Alternative perspectives. *NYS TESOL Journal, 1*(1), 2-10.

Gass, S. M., & Selinker, L. (2008). *Second language acquisition: An introductory course* (3rd edn). London: Routledge.

Georges, R. (1981). Do narrators really digress? A reconsideration of "audience asides" in narrating. *West. Folklore, 40*, 245-252.

Goldberg, A. E., & Casenhiser, D. (2008). Construction learning and second language acquisition. In P. Robinson & N. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 197-215). New York: Routledge.

Goldschneider, J., & DeKeyser, R. M. (2001). Explaining the 'natural order of L2 morpheme acquisition' in English: A meta-analysis of multiple determinants. *Language Learning, 51*(1), 1-50.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. Cowie (Ed.), *Phraseology: theory, analysis and applications* (pp. 145–160). Oxford: Oxford University Press.

Granger, S. (2012). Learner Corpora. In C. A. Chapelle (Ed.), *The encyclopedia of Applied Linguistics*. Oxford: Wiley-Blackwell.

Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2009). *International Corpus of Learner English* v2. Louvain-la-Neuve: Presses universitaires de Louvain.

Gries, S., & Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics, 9*(1), 97-129.

Groom, N. (2007). Phraseology and epistemology in Humanities writing. Unpublished PhD thesis, University of Birmingham.

Groom, N. (2010). Closed-class keywords and corpus-driven discourse analysis. In M. Bondi & M. Scott (Eds.), *Keyness in texts* (pp. 59-78). Amsterdam: John Benjamins.

Halliday, M. A. K. (1975). *Learning how to mean: Explorations in the development of language*. London: Edward Arnold.

Halliday, M. A. K. (1978). *Language as social semiotic*. London: Edward Arnold.

Halliday, M.A.K. (1994). *An introduction to functional grammar* (2nd edn). London: Edward Arnold.

Hanaoka, O., & Izumi, S. (2012). Noticing and uptake: Addressing pre-articulated covert problems in L2 writing. *Journal of Second Language Writing, 21*(4), 332-347.

Hinkel, E. (2011). What research on second language writing tells us and what it doesn't. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning*, Vol 2 (pp. 523-538). London: Routledge

Hoey, M. (2001). *Textual interaction: An introduction to written discourse analysis*. London: Routledge.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Abington: Routledge.

Hofstadter, D. R. (1999). *Godel, Escher, Bach: An eternal golden braid* (20th anniversary edn). New York, NY: Basic Books.

Housen, A. (2002). A corpus-based study of the L2-acquisition of the English verb system. In S. Granger, J. Hung, & S. Petch-Tyson (Eds), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 77-117). Amsterdam: John Benjamins.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30*(4), 461-473.

Howarth, P. (1996). *Phraseology in English academic writing: Some implications for language learning and dictionary making*. Tübingen: Max Niemeyer.

Howarth, P. (1998a). Phraseology and second language proficiency. *Applied Linguistics, 19*(1), 24-44.

Howarth, P. (1998b). The phraseology of learners' academic writing. In A. P. Cowie Ed.), *Phraseology: Theory, analysis, and applications* (pp. 161-186). Oxford: Oxford University Press.

Howatt, A. P. R. (2004). *A history of English language teaching* (2nd edn). Oxford: Oxford University Press.

Hu, X., & Liu, C. (2007). Restrictive relative clauses in English and Korean learners' second language Chinese. *Second Language Research, 23*, 263-187.

Huang, J., & Hatch, E. (1978). A Chinese child's acquisition of English. In E. Hatch (Ed.), *Second language acquisition: A book of readings* (pp. 118-131). Rowley, MA: Newbury House.

Huebner, T. (1979). Order-of-acquisition vs. dynamic paradigm: A comparison of method in interlanguage research. *TESOL quarterly, 12*(1), 21-28.

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Hunston, S. (2003). Lexis, wordform and complementation pattern: A corpus study. *Functions of Language, 10*(1), 31-60

Hunston, S. (2006). Starting with the small words: Patterns, lexis and semantic sequences. Paper given at Exploring the Lexis-Grammar Interface, University of Hannover, 6th October 2006. See also Hunston, S. (2008). Starting with the small words: Patterns, lexis and semantic sequences. *International Journal of Corpus Linguistics, 13*(3), 271-295.

Hunston, S. (2012). Corpus linguistics: Historical development. In C. A. Chapelle (Ed.), *The encyclopedia of Applied Linguistics*. Oxford: Wiley-Blackwell.

Hunston, S. (2013). Systemic functional linguistics, corpus linguistics, and the ideology of science. *Text & Talk,* 33, 617-640.

Hunston, S., & Francis, G. (1999). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: Benjamins.

Hyland, K., & Tse, P. (2005a). Evaluative that constructions: Signaling stance in research abstracts. *Functions of Language, 12* (1), 39-64.

Hyland, K. and Tse, P. (2005b). Hooking the reader: A corpus study of evaluative that in abstracts. *English for Specific Purposes 24*(1), 123-139.

Hymes, D. H. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds), *Sociolinguistics*. Harmondsworth, England: Penguin Books.

Isac, D., & Reiss, C. (2013). *I-language: An introduction to linguistics as cognitive science* (2nd edn). Oxford: Oxford University Press.

Ishikawa, S. (2011). A new horizon in learner corpus studies: The aim of the ICNALE project. In G. Weir, S. Ishikawa, & K. Poonpon (Eds.), *Corpora and language technologies in teaching, learning and research* (pp.3-11). Glasgow, UK: University of Strathclyde Publishing.

Jarvis, S., & Pavlenko, A. (2008). *Cross-linguistic influence in language and cognition*. London: Routledge.

Jenkins, J. (2013). *English as a lingua franca in the international university: The politics of academic English language policy*. Abington: Routledge.

Jensen, A. R. (1969). How much can we boost I.Q. and scholastic achievement? *Harvard EducationaI Review, 39*(1), 1-22.

Jespersen, O. (1924). *The philosophy of grammar*. London: Allen & Unwin.

Jespersen, O. (1909–49). *A modern English grammar on historical principles* (7 vols.). London: Allen & Unwin.

Kementerian Pendidikan Malaysia. (2000). *Sukatan Pelajaran Kurikulum Bersepadu Sekolah Menengah: Bahasa Inggeris*. Kuala Lumpur: Pusat Perkembangan Kurikulum.

Kennedy, G. (1998). *An introduction to corpus linguistics*. London: Longman.

Kjellmer, G. (1991). A mint of phrases. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics* (pp. 111-127). New York: Longman.

Klein, W. (1998). The contribution of second language acquisition research. *Language Learning, 48*(4), 527–550.

Klein, W., & Perdue, C. (Eds.) (1992). *Utterance structure: Developing grammars again*. Amsterdam: John Benjamins Publishing.

Klein, W., & Perdue, C. (1997). The Basic Variety (or: Couldn't natural languages be much simpler?). *Second language research, 13*(4), 301-347.

Kramsch, C. (2006). From communicative competence to symbolic competence. *Modern Language Journal, 90*(2), 249-252.

Krashen, S. (1977). Some issues relating to the Monitor Model. In H. Brown, C. Yorio & R. Crymes (Eds.), *On TESOL '77* (pp. 144-158). Washington, D.C.: TESOL.

Kress, G. (1994). Learning to write (2nd edn). London: Routledge.

Kroll, B. (Ed.). (1990). *Second language writing: Research insights for the classroom*. Cambridge: Cambridge University Press.

Labov, W. (1972). *Language in the inner city*. Philadelphia: University of Philadelphia Press.

Labov, W. & Harris, W. A. (1994). Addressing social issues through linguistic evidence. In J. Gibbons (Ed.), *Language and the law* (pp. 265–305). Harlow: Longman.

Lakshmanan, U., & Selinker, L. (2001). Analysing interlanguage: How do we know what learners know? *Second Language Research, 17*(4), 393-420.

Larsen-Freeman, D. (1976). An explanation for the morpheme acquisition order of second language learners. *Language Learning, 26*(1), 125-134.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics, 27*(4), 590-619.

Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, *30*(4), 579-589.

Larsen-Freeman, D., & Long, M. (1991). *An introduction to second language acquisition research*. London: Longman.

Latour, B. (1988). Mixing humans and nonhumans together: The sociology of a door-closer. *Social problems*, *35*(3), 298-310.

Leech, G. (1992). Corpora and theories of linguistic performance. In J. Svartvik (Ed.), *Directions in corpus linguistics: Proceedings of Nobel symposium 82* (pp. 125-148). Berlin: Mouton de Gruyter,

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus*. Harlow: Longman

Leki, I., Cumming, A., & Silva, T. (2008). *A synthesis of research on second language writing in English*. New York: Routledge.

Lindfors, J. W. (2008). *Children's language: Connecting reading, writing, and talk*. New York, NY: Teachers College Press.

Mackey, A. (1999). Input, interaction and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition, 21*(4), 557-587.

Marco, M. J. L. (2000). Collocational frameworks in medical research papers: A genre-based study. *English for specific purposes*, *19*(1), 63-86.

Matsuda, P. K., Ortmeier-Hooper, C., & Matsuda, A. (2009). The expansion of second language writing. In R. Beard, D. Myhill, J. Riley, & M. Nystrand (Eds.). *The SAGE handbook of writing development* (pp. 457-471). London: SAGE

Mauranen, A. (2012). *Exploring ELF: Academic English shaped by non-native speakers*. Cambridge: Cambridge University Press.

McCarthy, M. (1998). *Spoken language and applied linguistics*. Cambridge: Cambridge University Press.

McCarthy, M. (2001). Discourse. In R. Carter & D. Nunan (Eds.), *The Cambridge guide to teaching English to speakers of other languages*. Cambridge: Cambridge University Press.

McCarthy, M., & Carter, R. (2002). This that and the other: Multi-word clusters in spoken English as visible patterns of interaction. *Teanga: Yearbook of the Irish Association for Applied Linguistics, 21*, 30–52.

McCarthy, M., & O'Keefe, A. (2010). Historical perspective: What are corpora and how have they evolved? In A. O'Keefe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 3-13). London: Routledge.

McCarthy, M., O'Keeffe, A., & Walsh, S. (2010). *Vocabulary matrix*: *Understanding, learning, teaching*. Andover: Heinle, Cengage Learning.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice.* Cambridge: Cambridge University Press.

Meara, P. (1980). Vocabulary acquisition: a neglected aspect of language learning. *Language Teaching and Linguistics: Abstracts, 13*, 221-246.

Meara, P. (2002). The rediscovery of vocabulary. *Second Language Research, 18*(4), 393-407.

Meisel, J., Clahsen, H., & Pienemann, M. (1981). On determining developmental stages in natural second language acquisition. *Studies in Second Language Acquisition, 3*(2), 109-135.

Mel'čuk, I., Arbatchewsky-Jumarie, N., Elnitsky, L., Iordanskaja, L., & Lessard, A. (1984). *Dictionnaire explicatif et combinatoire du français contemporain: recherches lexico-sémantiques*, vol. i. Montréal: Les Presses de l'Université de Montréal.

Mel'čuk, I., Dagenais, L., Elnitsky, L, Iordanskaja, L., Lefebvre, M. N., & Mantha, S. (1988). *Dictionnaire explicatif et combinatoire du français contemporain: recherches lexico-sémantiques*, vol. ii. Montréal: Les Presses de l'Université de Montréal.

Milton, J. (2001). *Elements of a written interlanguage*. Hong Kong: Language Centre Research Reports v. 2.

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.

Milton, J., & Tsang, E. S. C. (1993). A corpus-based study of logical connectors in EFL students' writing: Directions for future research. In R. Pemberton & E. S. C. Tsang (Eds.), *Lexis in studies* (pp. 215-246). Hong Kong: Hong Kong University Press.

Mindt, I. (2011). *Adjective complementation: An empirical analysis of adjectives followed by that-clauses*. Amsterdam: John Benjamins.

Mitchell, R., Myles, F., & Marsden, E. (2013). *Second language learning theories* (3rd edn). Abington: Routledge.

Morgan, A., & Ziglio, E. (2007). Revitalising the evidence base for public health: An assets model. *Promotion & Education*, *14*(2 suppl), 17-22.

Musumeci, D. (2009). History of language teaching. In M. H. Long & C. J. Doughty (Eds), *The handbook of language teaching* (pp. 42-62). Malden, MA: Wiley-Blackwell.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle.

Nemser, W. (1971). Approximative systems of foreign-language learners. *International Review of Applied Linguistics, 9*(2), 115–123.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins.

Nunan, D. (1996). Issues in second language acquisition research: examining substance and procedure. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 349-374). London: Academic Press.

O'Rourke, B. & Ramallo, F. (2013). Competing ideologies of linguistic authority amongst "new speakers" in contemporary Galicia. *Language in Society, 42*(3), 287-305.

Ortega, L. (2009). *Understanding second language acquisition*. London: Hodder Arnold.

Ortega, L. (2013). SLA for the 21st century: Disciplinary progress, transdisciplinary relevance, and the bi/multilingual turn. *Language Learning*, *63*(s1), 1-24.

Ortega, L., & Byrnes, H. (2008). *The longitudinal study of advanced L2 capacities*. New York, NY: Routledge.

Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in SLA: Recent trends and future directions. *Annual Review of Applied Linguistics, 25*, 26-45.

Palmer, H. E. (1925). Conversation: The fundamental guiding principle for the student of conversation. *Bulletin, 10*, 2-5.

Palmer, H. E. (1933). *Second interim report on English collocations*. Tokyo: Institute for Research in English Teaching.

Phipps, A., & Gonzalez, M. (2004). *Modern languages: Learning and teaching in an intercultural field*. London: SAGE.

Pica, T. (2005). Second language acquisition research and applied linguistics. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 263-280). Mahwah, N.J.: Erlbaum.

Pike, K. (1981). Grammar versus reference in the analysis of discourse. In R. W. Bailey (Ed.), *Tagmemics, discourse, and verbal art* (pp. 47-64). Ann Arbor: University of Michigan Press.

Pienemann, M. (Ed.) (2005). Cross-linguistic aspects of processability theory. Amsterdam: John Benjamins.

Pienemann, M., Johnston, M., & Brindley, G. (1988). Constructing an acquisition-based procedure for second language assessment. *Studies in Second Language Acquisition, 10*(2), 217-243.

Polio, C. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing, 21*(4), 375-389.

Prodromou, L. (2003). In search of the successful user of English. *Modern English Teacher, 12*(2), 5-14.

Prodromou, L. (2008). *English as a lingua franca: A corpus-based analysis*. London: Continuum.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). A comprehensive grammar of the English language. London: Longman.

Rajagopalan, K. (2004). The philosophy of Applied Linguistics. In A. Davies & C. Elder (Eds.) *Handbook of Applied Linguistics* (pp. 397-420). New York: Blackwell Publishers.

Rampton, B. (1990). Displacing the 'native speaker': Expertise, affiliation, and inheritance. *ELT Journal*, *44*(2), 97-101.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Renouf, A., & Sinclair, J. (1991). Collocational frameworks in English. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics: Studies in honour of Ian Svartvik* (pp. 128—143). London: Longman.

Rutherford, W. (1988). *Second language grammar: learning and teaching*. London: Longman.

Schachter, J. (1974). An error in error analysis. *Language Learning, 24*(2), 205-214.

Schmitt, N. (Ed.) (2004). *Formulaic sequences*: Acquisition, processing and use. Amsterdam: John Benjamins.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke: Palgrave Macmillan.

Schmitt, N., & McCarthy, M. (Eds.) (1997). *Vocabulary: Description, acquisition and pedagogy*. Cambridge: Cambridge University Press.

Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education*. Amsterdam: John Benjamins.

Seidlhofer, B. (2011). *Understanding English as a lingua franca*. Oxford: Oxford University Press.

Seidlhofer, B. (2012). Corpora and English as a lingua franca. In K. Hyland, M. H. Chau & M. Handford (Eds.), *Corpus applications in Applied Linguistics* (pp. 135-149). London: Continuum.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching*, *10*(4), 209-232.

Shanker, S. G., & Taylor, T. J. (2001). The house that Bruner built. In D. Bakhurst & S. G. Shanker (Eds.), *Jerome Bruner: Language, culture, self* (pp. 50-70). London: SAGE.

Sinclair, J. M. (1966). Beginning the study of lexis. In C.E. Bazell, J.C. Catford, M.A.K. Halliday & R.H. Robins (Eds.), *In memory of J.R. Firth* (pp. 410-430). London: Longman.

Sinclair, J. M. (Ed.). (1987). *Looking up: An account of the COBUILD project in lexical computing and the development of the Collins COBUILD English language dictionary*. London: HarperCollins.

Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, J. M. (1992). Trust the text. In L. Ravelli & M. Davies (Eds), *Advances in systemic-functional linguistics: Recent theory and practice* (pp. 1–19). London: Pinter.

Sinclair, J. M. (1996). The search for units of meaning. *Textus, 9*(1), 75-106.

Sinclair, J. M. (2004). *Trust the text: Language, corpus and discourse*. London: Routledge.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Spada, N., & Lightbown, P. M. (2002). Second language acquisition. In N. Schmitt (Ed.), *An introduction to Applied Linguistics* (pp. 115-132). London: Arnold.

Spivak, G. (1999). Translation as culture. In I. Carrera Suarez, A. Garcia Ferandez, & M. S. Suarez Lafuente (Eds), *Translating cultures* (pp. 17–30). Hebden Bridge: Dangaroo Press.

Stefanowitsch, A., & Gries, S. (2003). Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics, 8*(2), 209-243.

Stubbs, M. (2004). Language corpora. In A. Davies & C. Elder (Eds.), *Handbook of Applied Linguistics* (pp. 106-32). Oxford: Blackwell.

Stubbs, M., & Barth, I. (2003). Using recurrent phrases as text-type. *Functions of language*, *10*(1), 61-104.

Swain, M. (2006). Languaging, agency and collaboration in advanced second language learning. In H. Byrnes (Ed.), *Advanced language learning: The contributions of Halliday and Vygotsky* (pp. 95–108). London: Continuum.

Thomas, G. E. (1983). The deficit, difference, and bicultural theories of black dialect and nonstandard English. *The Urban Review*, *15*(2), 107-118.

Thompson, G., & Hunston, S. (Eds.). *System and corpus: Exploring connections*. London: Equinox.

Thorndike, E. L., & Lorge, I. (1944). *A teacher's word book of 30,000 words*. New York: Teachers' College, Columbia University.

Timmis, I. (2002). Native-speaker norms and International English: A classroom view. *ELT Journal, 56*(3), 240–9.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins.

Tomasello, M. (2003). *Constructing a language.* Boston, MA: Harvard University Press.

Tono, Y. (2004). Multiple comparisons of IL, L1 and TL corpora: The case of L2 acquisition of verb subcategorization patterns by Japanese learners of English. In G. Aston, S. Bernardini & D. Stewart (Eds), *Corpora and language learners* (pp. 45-66). Amsterdam: John Benjamins.

VanPatten, B., & Williams, J. (Eds.) (2007). *Theories in second language acquisition: An introduction*. Mahwah, NJ: Lawrence Erlbaum.

Wells, G. (2009). *The meaning makers: Learning to talk and talking to learn* (2nd edn). Bristol: Multilingual Matters.

West, M. (1953). *A general service list of English words*. London: Longman.

White, L. (2003). *Second language acquisition and Universal Grammar*. Cambridge: Cambridge University Press.

White, L. (2013). Universal Grammar (UG) and SLA. In P. Robinson (Ed.), *The Routledge encyclopedia of second language acquisition* (pp. 670-674). New York, NY: Routledge.

Wigglesworth, G., & Storch, N. (2012). What role for collaboration in writing and writing feedback. *Journal of Second Language Writing, 21*(4), 364-374.

Wiktorsson, M. (2000). Prefabricated phrases in learner language. A corpus-based study comparing advanced EFL writing with native English writing. In G. Byrman, H. Lindquist & M. Levin (Eds.), *Corpora in research and teaching* (pp. 304-314). Uppsala: ASLA/FUMS.

Wiktorsson, M. (2001). Register differences between prefabs in native and EFL English. In S. Manninen & C. Paradis (Eds), *Working papers in linguistics* (Vol 1) (pp. 85-94). Lund: Lund University.

Williams, J. (2012). The potential role(s) of writing in second language development. *Journal of Second Language Writing, 21*(4), 321-331.

Wittgenstein, L. (1953). *Philosophical investigations*. G.E.M. Anscombe (Trans.). G.E.M. Anscombe & R. Rhees (Eds.). Oxford: Blackwell.

Worlock Pope, C. (Ed.) (2010). The Bootcamp discourse and beyond. Special issue of the *International Journal of Corpus Linguistics, 15*(3).

Žižek, S. (2006). *The parallax view*. Cambridge, MA: MIT Press.

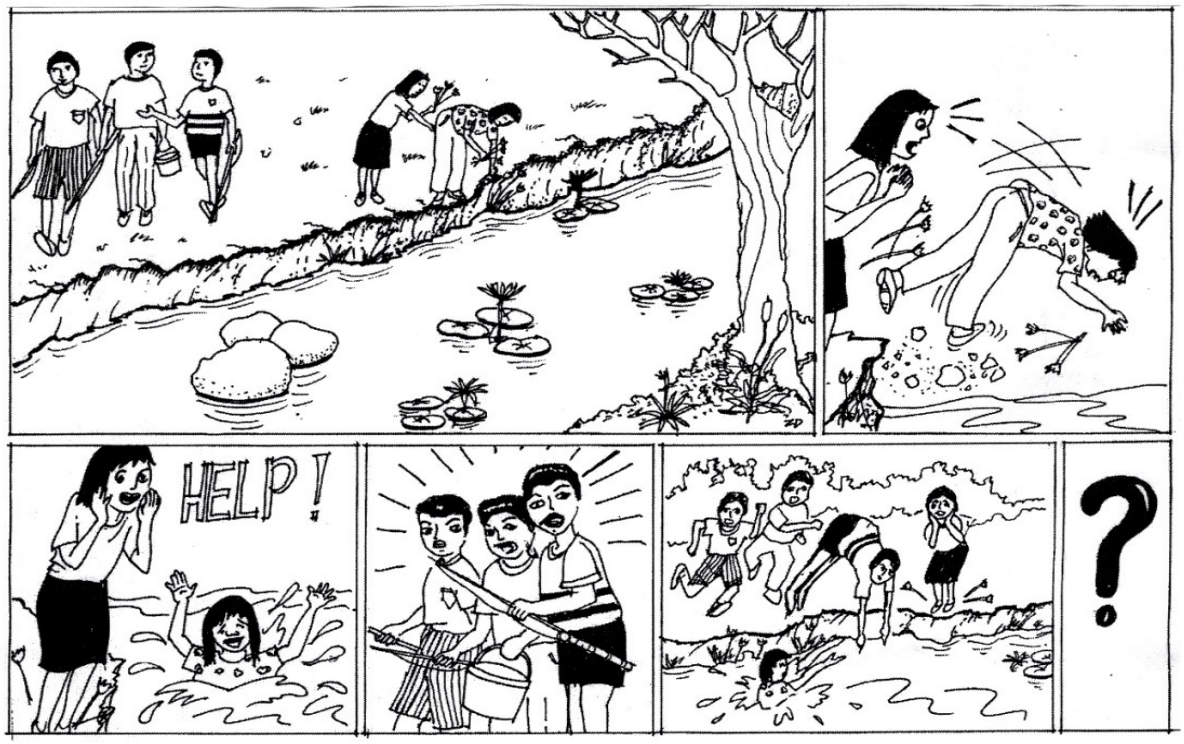
Zobl, H. (1980). The formal and developmental selectivity of L1 influence on L2 acquisition. *Language Learning, 30*(1), 43-57.

Zobl, H. (1982). A direction for contrastive analysis: The comparative study of developmental sequences. *TESOL Quarterly, 16*(2), 169-183.

.

APPENDIX

In not fewer than 100 words, write a story based on the series of pictures below. Make your story as interesting as possible.



[Source: Abd. Samad et al., 2002]