Aafke Buyl
Vrije Universiteit Brussel
Faculteit der Letteren en Wijsbegeerte
Studiegebied Taal- en Letterkunde
Promotor: Prof. dr. A. Housen

The Development of Receptive Grammar Knowledge in English as a Second Language:

A Cross-Sectional Study

"Proeve ingediend voor het behalen van de graad van:
Master na Master of Advanced Studies in Linguistics"

Academiejaar 2009-2010

Grant Agreement Number: 2008 - 3378 / 001 - 001

Summary

The Development of Receptive Grammar Knowledge
in English as a Second Language:
A Cross-Sectional Study

Aafke Buyl
Supervisor: Prof. dr. A. Housen
Vrije Universiteit Brussel
2009-2010

Aim and Methods

This study aims at gaining insight in the dynamics of receptive second language (L2) grammar acquisition. It is investigated whether the acquisition of receptive L2 grammar knowledge contains developmental patterns. In a cross-sectional study, a receptive grammar test containing nine grammatical phenomena is administered to participants learning English as an L2.

Because the participants are children – aged 7 to 10 – who acquired their L2 in immersion education, this study may also contribute to the debates about the age factor in SLA and about bilingual education.

The grammar test (GT) used was developed by ELIAS, a Comenius-funded research project investigating immersion education in Europe.

Results

To date scholars investigating L2 grammar acquisition have focused almost exclusively on the development of productive language knowledge. A literature review shows that it is often (tacitly) assumed that the developmental patterns attested in productive grammar development also pertain to receptive grammar development. However there is very little empirical evidence to support this. Neither the morpheme studies conducted in the 1970s nor the interlanguage studies of grammatical subsystems conducted over the course of several decennia have investigated receptive language data.

Rather, the extrapolation of results from production to reception appears to have issued from the assumption that language features must be comprehended before they can be acquired for production, as well as on informal observations that the acquisition of receptive skills precedes the acquisition of productive skills – an assumption which has led to the traditional view, in Second Language Acquisition (SLA), of receptive language development as being necessary for and hence preceding productive grammar acquisition.

Some scholars have rejected this traditional view by pointing to studies which show that the acquisition of productive language skills can precede that of receptive language knowledge. Because not all scholars accept the evidence for this primacy of production over reception, however, it is not clear whether the traditional view on language reception can be rejected.

The aim of this study therefore is to respond to this lacuna in SLA research. By conducting a receptive grammar test (GT) it is investigated whether developmental patterns exist in receptive grammar acquisition, and if so whether these patterns are in any way similar to the ones found in studies of productive grammar acquisition. The GT used contains nine grammatical categories: canonical word order (SVO), sentence negation (NEG), personal pronouns in the 3rd person singular, in the nominative case (PROsubj) and in the accusative case (PROobj), possessive determiner (POSS), genitive marker *–'s* (GEN), plural marker *–s* (PLU), morphology of copular verbs (COP) and morphology of full verbs (VERB) – in both cases of the 3rd person singular and plural in the simple present.

To the extent that this is possible in a study of this scope, it is examined whether the explanations that have been offered for developmental patterns in productive grammar acquisition might pertain to receptive acquisition. Examples of verifiable explanations that scholars have posited to account for productive grammar acquisition are frequency of occurrence, markedness or, in Pienemann's Processability Theory, mental processing constraints.

The GT used in this study was developed by ELIAS to investigate the acquisition of English as an L2 in early immersion education across Europe. Likewise the data examined in this study were collected, within the wider frame of the ELIAS study, in an early English immersion school in the French Community in Belgium. Hence the results of this study may also contribute to the debates about the age factor in SLA and about processes and outcomes in bilingual education.

Finally, because the pattern of receptive grammar knowledge is relatively unexplored territory, as yet no research methodologies or test materials exist which have shown their value for the investigation of receptive L2 grammar acquisition through use in a large number of studies. With this in mind, considerable attention is paid in this study to the advantages and disadvantages of the grammar test used.

By rank ordering grammatical categories for difficulty and by using implicational scaling, it was found that the nine categories in this study appear to be acquired in an implicational order. This implicational acquisition order can be identified as follows: 1 NEG and SVO, 2 GEN, 3 PLU, and 4 POSS, COP, PROobj, VERB and PROsubj. Within each rank, differences between the categories are not significant.

Since this rank order is different from the rank orders identified in studies of productive L2 grammar acquisition, it is unlikely that explanations such as frequency of occurrence and markedness also apply to the acquisition order found in this study. Testing Processability Theory proved difficult because not all 'processing procedures' were present in the GT used.

Importantly, several factors were identified which warn against interpreting the results from this tentative study as conclusive evidence – among others the incompleteness of the implicational scales and the semantic processing difficulties encountered by the participants.

Furthermore, the findings should not be generalized too readily. Results may be different for other grammatical categories, for participants learning the L2 in other contexts and for older L2 learners. Similarly, although no conspicuous L1 effects were identified in this study, it is possible that L1 background affects the route of receptive grammar development.

Therefore the challenge for future researchers investigating receptive grammar acquisition lies in developing tests and methodologies which control for semantic comprehension, as well as in conducting research in a wide variety of settings so that the influence of factors such as age, learning setting and L1 background can be assessed.

Finally, a cross-sectional analysis of the test results revealed the acquisition of receptive grammar knowledge to proceed relatively slowly, suggesting that young L2 learners acquire the L2 at a slower rate than older L2 learners, and/or that grammar acquisition in immersion education is a slow process. However, a comparison with older learners and with learners in other settings is required for more conclusive evidence.

Abstract

The Development of Receptive Grammar Knowledge
in English as a Second Language:
A Cross-Sectional Study

Aafke Buyl
Supervisor: Prof. dr. A. Housen
Vrije Universiteit Brussel
2009-2010

The aim of this study is to gain insight into receptive L2 grammar acquisition, an as yet unexplored field in SLA. It is investigated whether receptive L2 grammar acquisition, like its productive counterpart, proceeds in a relatively fixed manner. To this end a grammar test containing nine grammatical categories has been administered to francophone children (aged 7 to 10) learning English in immersion education. The results of this exploratory study point to the existence of a developmental order, but one which diverges from the pattern found in productive grammar acquisition. In addition, a cross-sectional analysis of three groups of participants with different amounts of L2 contact (18, 29 or 41 months) reveals a slow rate of acquisition. However several factors were identified which warn against extrapolating the tentative findings from this study or interpreting them as conclusive evidence. Future research will have to corroborate or contradict the results.

Key Words

Second Language Acquisition; English as a Second Language; Morphosyntax; Grammar Acquisition; Receptive Language Knowledge; Immersion Education

Trefwoorden

Tweedetaalverwerving; Engels als tweede taal, Morphosyntaxis; Spraakkunstverwerving; Receptieve Taalkennis; Immersieonderwijs

Acknowledgements

I would like to offer my sincere thanks to prof. dr. A. Housen, my supervisor, for his advice and support, and for offering me the chance to cooperate in the ELIAS study, which has been a very enriching experience.

Thanks are also due the ELIAS team, for the use of their grammar test and test results and for their patience in the last weeks before the deadline of this ManaMa thesis.

Finally, I would like to thank the staff of the Ecole Cheval Bayard, and in particular the primary school teachers - Claudia, Barbara and Sophie - for their relentless enthusiasm and for their quick acceptance of my presence in their classrooms. And of course this study would not have been possible without the pupils' cooperation in my English 'games'.

This study has been funded with support from the European Commission. This publication [communication] reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Contents

Abbreviations

BSM	*Bilingual Syntax Measure*
CAH	*Contrastive Analysis Hypothesis*
CA	*Contrastive Analysis*
EA	*Error Analysis*
IL	*Interlanguage*
IP	*Input Processing*
L1	*First Language*
L2	*Second Language*
NPAH	*Noun Phrase Accessibility Hierarchy*
OT	*Optimality* Theory
P1	*L2 Contact Group: 1st Year of Primary School; L2 Contact: 18 Months*
P2	*L2 Contact Group: 2nd Year of Primary School; L2 Contact: 29 Months*
P3	*L2 Contact Group: 3rd Year of Primary School; L2 Contact: 41 Months*
PT	*Processability Theory*
SLA	*Second Language Acquisition*
SOC	*Suppliance in Obligatory Occasion*
TL	*Target Language*

Tables

Figures

0 Introduction

This study deals with the acquisition of receptive knowledge of second language (L2) grammar, as contrasting with the acquisition of productive grammar knowledge.

Although knowing a language entails knowing both how to produce and how to comprehend utterances in that language, research to date on Second Language Acquisition (SLA) has mainly focused on the acquisition of *productive* knowledge of L2 grammar. The development of *receptive* L2 grammar knowledge, by contrast, is an area several researchers have declared an unexplored territory in SLA research, and is one of the major challenges which future SLA research needs to respond to (Ellis, 2008). To redress this imbalance, this study will explore the dynamics of receptive grammar development by L2 learners of English.

Studies of *productive* grammar development have demonstrated that a relatively fixed and possibly universal chronology can be discerned in the development of productive L2 grammar: L2 learners have been found to consistently acquire certain grammatical features of the L2 before others and to furthermore pass through a fixed sequence of developmental stages in their development of individual grammatical features. It has also been suggested that these fixed patterns of grammar development manifest themselves independently of the learners' L1, age or of the L2 learning context. In view of the vast body of research which points to the existence of these developmental patterns in productive L2 grammar acquisition, this study particularly seeks to gain insight into whether or not developmental regularities can be found in receptive L2 grammar development.

1 Early SLA and the Focus on Language Production

Research into the acquisition of L2 grammar has a fairly recent history. This section will briefly describe the background against which this interest originated. In addition, it will be made clear that the lack of attention for language reception has been a constant factor in SLA from the earliest grammar acquisition studies onwards.

SLA originated in the 1960s. In the 1950s the prevailing view on language learning was behaviourism, which saw L1 and L2 learning as the development of distinct sets of

habits (Lightbown & Spada, 1993). Within this behaviourist framework, the Contrastive Analysis Hypothesis (CAH) claimed that errors occur in the L2 where the L1 and the L2 are dissimilar; where they are similar, the learner can rely on the L1 to construct the L2. Thus transfer was a key concept at the time. In a strong version, the Contrastive Analysis (CA) method compared the L1 and L2 to predict errors. In a weak version it was used to explain errors. CA, further, was not intended to gain insight into L2 acquisition as such but was meant to improve language pedagogy (Hatch, 1983; Long and Sato, 1984).

Influenced by Chomsky's mentalist views on L1 acquisition, the behaviourist notion of L2 learning was largely abandoned in the 1960s (Cook, 1993). In Corder's (1981) words, 'it was natural that one should ask whether the cognitive processes which came into play in L1 acquisition were the same as those used in L2 learning' (p. 2). According to mentalist views on L2 learning, moreover, transfer was non-existent, thus rendering the CAH void (Gass & Selinker, 2001).

In CA, furthermore, errors were seen as 'a product of imperfect learning' (Gass & Selinker, 2001, p. 102). However, incited by Corder's *The Significance of Learners' Errors* (1967), this perspective was renounced. In his article Corder (1967) revalued errors for a number of reasons. The most important for the field of SLA are that 'we can regard the making of errors as a device the learner uses in order to learn' (p. 167) and, secondly, that errors provide the researcher with insight into the process of L2 acquisition. Therefore scholars agree that Corder's work formed a breakthrough in L2 acquisition research (e.g. Selinker, 1992).

The preferred research method in the 1960s, then, was Error Analyses (EA). In EA, L2 learners' errors are compared with the TL forms so as to gain insight in the L2 acquisition process and to find evidence of systematicity in L2 development (Corder, 1967; Long & Sato, 1984). Despite the criticisms directed towards the methodology of EA in later years and the consequent demise of EA, it is widely acknowledge that this methodology has played an important part in the development of the field of SLA. In addition, EA is of particular interest for this study because the findings of EA contained signs of systematicity in learner language (Ellis, 2008; Long and Sato, 1984). Although EA did not yet 'succeed in providing clear and conclusive evidence of developmental

patterns' (Ellis, 2008, p. 68) it may certainly have contributed to the growing interest in the phenomenon of systematicity.

In short, the shift form behaviourist to mentalist views along with the changing perspective on errors are considered crucial factors in the development of SLA, which in turn has been a sine qua non for the study of L2 grammar acquisition (Gass & Selinker, 2008; Selinker, 1992). Interesting to note, moreover, is that from the first L2 acquisition studies onwards production data has served as the sole source for gaining insight into this complex matter. Furthermore, systematicity in learner language has been a constant factor in SLA research from the beginning. Therefore this will also be a central factor in this study.

2 Main Concepts and Terms

A distinction will be made in this study between *Second Language Acquisition* (*SLA*), the field of research, and *L2 acquisition*, the process of acquiring a second language. *Second* will thereby refer to any language other than the L1.

Also in line with contemporary SLA research a distinction will be made between *order of acquisition* and *sequence of acquisition* (Ellis, 2008). The *order of acquisition* has to do with the acquisition of TL features one after another in a relatively fixed and possibly universal order. *Sequence of acquisition* refers to the different stages which can be observed within the acquisition of one particular TL feature. Both *order* and *sequence* of acquisition are encompassed in the terms *developmental patterns, systematicity or developmental regularities.*

Finally, a term that is frequently used in SLA research is *interlanguage*. As Selinker (1992) explains, in Contrastive Analysis (see above) 'two linguistic systems were talked about, N[ative] L[anguage] and TL (…), but *three* systems, with the third unnamed, seemed to me always to be implied' (p. 222). IL, then, refers to an L2 learner's 'interim competence that contains elements from both the L1 and the L2 grammar, but also elements that go beyond both' (Ortega, 2009a, p. 81; see also Selinker, 1967). A further discussion of the notion of IL would lead this study too far. For the sake of this study, it suffices to say that the term *IL* is now commonly used in SLA to refer to the mental grammar of L2 learners at one particular moment in their development (Ellis, 2008;

White, 2005), to ‘the series of interlocking systems which characterize acquisition’, as well as to particular L1/L2 combinations (Ellis, 2008, p. 968).

3 Overview

The main body of this study consists of two chapters. In the first chapter some of the relevant literature concerned with L2 grammar acquisition will be presented. Attention will be paid here to both receptive and productive grammar acquisition. On the one hand, it will be examined how language comprehension and the acquisition of receptive language skills are perceived of in SLA, in particular with regard to its relation to language production. On the other hand, the large body of research concerned with productive L2 grammar acquisition – more specifically the systematicity attested in this acquisition process – will be discussed. Given the amount of research which has investigated productive grammar acquisition, this discussion will inevitably take up some space.

In the second chapter an empirical investigation into the acquisition of receptive knowledge of L2 English grammar is presented. In this investigation a receptive grammar test has been used to assess the knowledge of nine – predominantly morphological – grammatical categories. Considering the relatively unexplored nature of receptive grammar acquisition, this study should not be seen as an attempt to bring forth conclusive evidence with regard to the existence of developmental patterns in receptive grammar acquisition or its relationship to language production. Rather it is an exploratory study which hopes to help broadening the scope of L2 grammar acquisition research.

Because the participants in this study are young L2 learners of English (age of onset 5 to 6 years, L2 contact between 19 and 41 months) who are furthermore enrolled in English immersion education, this study may also contribute to the scholarly debates about age of acquisition on the one hand and bilingual education on the other hand.

I

Literature Review: The Acquisition of Second Language Grammar Knowledge

0 Introduction

In this first chapter, some of the scholarly literature on receptive and productive L2 grammar acquisition is dealt with.

Section I explores the basic dynamics of language comprehension and seeks to gain insight into the relationship between productive and receptive language acquisition. The imbalance in the use of productive versus receptive language data in SLA research may to some extent be attributed to the fact that receptive knowledge of a language consists of 'a "corpus" of mental representations' (Berthele, Kaiser & Peyer, in press), the product of which is much less easily accessible than the product of productive language skills – namely language production data. However, a look at the literature on L1 and L2 acquisition in general, and on grammar acquisition in particular, reveals that there is more behind this imbalance than the difficulties involved in studying receptive language knowledge.

Sections 2 and 3 give an overview of some of the most relevant literature concerned with developmental patterns in L2 grammar acquisition. In order to be able to determine whether the acquisition of receptive L2 grammar is likely to resemble productive L2 grammar acquisition, it is necessary to have an understanding not just of the relationship between language production and language reception but also of how the acquisition of productive L2 grammar knowledge is believed to proceed. Therefore section 2 summarises the most important findings of empirical research into productive L2 grammar acquisition. In keeping with the research-then-theory approach that has long been taken in this branch of SLA, section 3 shifts to focus to the explanations that have been offered for developmental patterns.

To conclude this literature review, section 4 will discuss L1 background, age of acquisition and learning settings – three major factors which have been claimed not to affect the route of L2 grammar development.

1 Language Decoding and Receptive Language Acquisition

1.0 Introduction

Language reception can be studied from two perspectives. A first, psycholinguistic perspective examines the processing of oral and written input in the human mind. This mental activity, also referred to by term *decoding*, has been extensively studied (Aitchison, 1998). Because this process is not unimportant for the cross-sectional study in chapter 2, the basic dynamics of receptive language processing will be presented in section 1.

Language reception can also be studied within the field of SLA. In this case, scholars are interested in the processes involved in the acquisition of receptive L2 knowledge (e.g. grammar, vocabulary, pragmatics). Unlike psycholinguistic language decoding, the acquisition of receptive language knowledge has hitherto been largely ignored by SLA research. Berthele, Kaiser and Peyer (in press) appeal to the hidden and hence opaque nature of receptive language knowledge as a (partial) explanation for the lack of attention for receptive L2 acquisition. However, this hidden nature did not keep psycholinguists from exploring the decoding of language in the human mind. Section two therefore will examine the relevant SLA literature, with the aim of gaining insight into the relationship between language production and reception and furthermore understanding why receptive language acquisition has been consistently ignored by scholars.

1.1 Receptive Language Processing or 'Decoding'

From a psycholinguistic or cognitive perspective, ample attention has been paid to the processes involved in language reception (or *decoding*, as opposed to *encoding* or language production). According to Aitchison (1998), language reception has even been studied 'more intensively' than language production. Although many questions are still open with regard to how exactly listeners go about decoding language, while moreover much of what researchers believe to know about comprehension can – due to the hidden nature of this process –be nothing than mere speculation, language comprehension is nonetheless better understood then language production (Aitchison, 1998).

Language comprehension is not a passive process of registering input. Rather listeners 'actively reconstruct both the sound and syntax of an utterance in accordance with their expectations' (Aitchison, 1998, p. 201). There are several ways in which this can be done.

In an inferential encoding process, Gass & Selinker (2008) explain, listeners for example rely on contextual clues and prior knowledge to achieve a general understanding of the message – that is, an understanding on a semantic level only. Similarly, Aitchison describes how in 'bottom-up processing' listeners may 'gather particular words together and try to make sense of them' (Aitchison, 1998, p. 236).

In a seemingly contrastive type of input processing, listeners pay attention to form (i.e. the grammatical structure of the input). They thereby rely on a structural analysis of the input, while also using their expectations about the syntactic form of sentences (e.g. SVO as the canonical word order) to understand speech or written input (Gass & Selinker, 2008). In the latter case, a 'more analytic understanding' takes place (ibid., p. 484). Importantly, not just syntax but also the lexical information contained in the message plays an important part in this analytic comprehension, for example in helping listeners to work out syntactic relationships.

Although semantic and grammatical processing are sometimes viewed as contrastive (Aitchison, 1998), Gass & Selinker explain that there are different levels of comprehension, situated on a continuum between semantic understanding on the one extreme and analytic understanding on the other extreme.

Language comprehension has also been studied within the field of SLA. That is to say, theories have been developed to account for how learners of an L2 comprehend the TL. VanPatten's Input Processing (IP) model (VanPatten, 2002) or the emergentist Competition Model (Ellis, 2008) are just two examples of such theories. To illustrate, according to VanPatten's IP model, learners have a limited processing capacity, so that they cannot attend to different stimuli at the same time. Consequently, in processing input, learners will first process input for meaning rather than for form. To do so they will, among other things, pay more attention to content words than to grammatical items. Hence, it is posited that processing at a semantic level – one end of the above-mentioned continuum – plays a more important role than analytic processing when processing a not yet fully acquired L2.

Although the psycholinguistic process of language decoding is not the central topic of this study, the above discussion is nonetheless within the scope of this study since the empirical study discussed in chapter II necessitates a basic understanding of language decoding – as will become clear in the chapter in question.

1.2 Receptive (and Productive) Language Acquisition

Far less frequently studied than the mental processes involved in language comprehension is the acquisition of receptive language skills – be it in L1 or L2 acquisition.

With regard to L1 and L2 acquisition in general, scholars note that it has long been thought – both in folk wisdom and in scholarly opinion – that comprehension always precedes production, because 'before learners start a new task they need to understand what they are doing' (Tasseva-Kurktchieva, 2008, p. 242). Furthermore, as Keenan and MacWhinney (1987) observe 'comprehension is viewed as the primary source of learning to produce language' (p. 149).

In SLA research concerned with the acquisition of grammar, Krashen indeed argued that 'comprehensible input' is sufficient for language acquisition to take place (e.g. Krashen, 1981; 1985). However scholars now agree that in order for learning to take place from input, input alone is not sufficient but rather must be combined with output. Furthermore, scholars now also underscore the difference between comprehension and acquisition. Since comprehension is merely the 'processing [of] language to get the message' (Cook, 1996, as cited in Gass & Selinker, 2008, p. 376), it 'will not necessarily lead to grammar formation' (Gass & Selinker, 2008, p. 486).

Rather, in order for language acquisition to take place learners must process input in such a way that they 'break the code'. Put differently, they must have an analytic understanding of the input in order that input can turn into *intake*. This intake can then be used to acquire the grammar of the language. In short, 'semantic comprehension is a prerequisite to syntactic comprehension, and syntactic comprehension is a prerequisite to acquisition. Neither guarantees the following step' (ibid., p. 375).

In short, although learning a new language entails the development of both receptive and productive language skills, researchers have rarely focus on the development of receptive skills as such. Rather, comprehension – more particularly analytic

comprehension – is viewed as a requirement for acquisition to take place. It follows from this that, before L2 learners (or L1 learners, for that matter) are able to produce a grammar feature, they have acquired a receptive knowledge of this feature.

Once the TL grammar has been acquired, traditional views on grammar see this grammar as a 'direction-insensitive system of rules' (Hendriks, 2007, p. 240). That is to say, people have one mental grammar which enables them to both comprehend and produce language. This belief is supported by the symmetry between a person's comprehension and production competences, as can be found in fully acquired L1s in particular: when a person is able to produce a grammar structure, s/he is also able to comprehend it, and vice versa (Hendriks, de Hoop & Lamers, 2005; Hendriks & Koster, in press).

Different language theories of course have different views on the relative importance of input in L2 acquisition, and on the specific ways in which learners use this input to acquire language. Universal Grammar, for example, views input as a trigger which activates humans' innate, mental grammar (Ellis, 2008). In other, cognitive accounts of language acquisition, the existence of an innate grammar system is rejected and learners are said to use general learning mechanisms for acquiring the TL grammar (ibid.). Furthermore the form which mental grammars take in language theories varies from one theory from another. In spite of these differences, however, SLA researchers, across all theories and views on language acquisition, seem to have adopted, first, the idea that language comprehension serves acquisition and hence precedes production, and secondly, the position that once the TL is acquired knowledge of the language (or mental grammar system) is used for both comprehension and production.

The above explains why language reception has not been studied more frequently. Although it may be true, as Berthele et al. write, that the product of receptive language knowledge is much less easily accessible, it can be added to this that there was also no real need to study the acquisition of receptive language skills. Based on the observation that learners understand a new language before they produce it and on the assumption that reception serves acquisition, it appears logical that whatever observations are made on the basis of productive language data (as in section 2 below) will also apply to language reception.

Recently, however, the traditional views on the relationship between receptive and productive second language acquisition have been challenged. Although, as de Hoop and Krämer (2006) write with regard to L1 acquisition, 'in general (…) young children's abilities in comprehension exceed their abilities in production' (p. 120), several cases have been attested where production precedes comprehension. These cases concern not just L1 acquisition (e.g. Hurewitz, Brown-Schmidt, Thorpe, Gleitman & Trueswell, 2000; Spenader, Smits, & Hendriks, 2009) but also L2 acquisition (Unsworth, 2007; Tasseva-Kurktchieva, 2008). Furthermore, they relate to both the acquisition of grammar and the acquisition of lexical items (Bates, Dale, & Thal, 1995).

According to some, the results from studies which have shown the primacy of productive acquisition over receptive acquisition may not be valid due to methodological flaws (Tasseva-Kurktchieva, 2008; Unsworth, 2007). Bates et al. (1995), for example, argue that in studies where the productive acquisition of grammatical items preceded the receptive acquisition, the results were due either to the 'complex tasks demands that obscure the child's actual knowledge of grammatical structures' (p. 8), or to the fact that the children produced unanalyzed and memorized chunks of language which contained the grammatical items being investigated, thereby giving the impression that they had acquired these grammatical features while in reality they had not.

Despite these criticisms, other researchers have taken this evidence serious. In trying to explain the primacy of language reception over language production, some scholars have appealed to processing-problems or computational limitations encountered by learners in the course of the comprehension process (e.g. Hurewitz et al., 2000; Unsworth, 2007). Still others invoke the absence of specific pragmatic knowledge required for receptive processing which is not required for productive processing – though this view appears to be disputed (for an overview, see Hendriks & Coster, 2010).

Thirdly, some scholars regard these production/comprehension asymmetries as 'a result of the grammar'. While the possibility that learners have different grammars for comprehension and production is generally rejected (e.g. Hendriks & Koster, in press; Tasseva-Kurktchieva, 2008), the traditional direction-insensitive view on grammar is questioned by some (e.g. Hendriks, 2007). Notably Optimality Theory (OT), which

posits a direction-sensitive grammar, accounts for asymmetries between production and comprehension. In Hendriks' (2007) words, 'in Optimality Theory (OT), production and comprehension are modeled as different directions of use of the same grammar' (p. 2). It is possible, therefore, that a grammar feature can be mastered in one direction but not in another. Thus, OT accounts for the usually found primacy of language reception over language production as well as for the less usual primacy of language production over language reception. For more information on Optimality Theory, the reader is referred to Hendriks (2007).

1.3 Conclusion

This section has dealt both with the psycholinguistic process of encoding comprehension and with the acquisition of receptive and productive language skills in the acquisition of L1 grammar.

With regard to decoding, it has been explained that according to the psycholinguistic literature there are two major ways in which listeners can process input: either the general semantic content only can be processed or the input can be processed at an analytic level. In reality, these two seemingly contrastive ways of decoding constitute the two extremes of a continuum.

Concerning the acquisition of receptive and productive skills, the above discussion has shown that the relationship between the acquisition of receptive grammar knowledge and the acquisition of productive language skills is far from straightforward. In traditional views, receptive knowledge is seen as preceding productive skills in the acquisition of L2 grammar – and other aspects of language for that matter – because it is believed that in order for acquisition to take place learners must first process input at an analytic level. On the basis of this analytic processing, it is posited, input can turn into intake, which in turn is used for acquisition. Furthermore, SLA researchers usually regard grammar as a uni-directional system: once the TL grammar system has been acquired, it is used both for comprehending and producing language.

On the basis of fairly recent empirical evidence which shows a primacy of language production over language reception, however, scholars have challenged these traditional views. Although to date the traditional views on language reception still dominate SLA, and scholars have even challenged the validity of studies claiming to contradict this

view, it may be clear that more research will be needed to either reject or confirm the more recent theories.

In traditional views on language reception, studies of productive language acquisition can also be taken to apply to language reception – as has frequently, though usually tacitly, been done (see section 2 below). The recent challenges that have been posed to the traditional views make that these claims can be questioned.

2 The Route of L2 Grammar Acquisition: Empirical Evidence

2.0 Introduction

Several decades of SLA research have brought forward evidence that L2 grammar acquisition is marked by developmental regularities, in the form of developmental orders and developmental sequences (see *Introduction: Main Concepts and Terms*). This section will start where the overview of early SLA in the introduction (see *Introduction*: *Early SLA and the Focus on Language Production*) set off, namely in the early years of SLA after the heyday of Error Analysis, as this is when the first conclusive signs of developmental patterns in productive grammar acquisition were attested.

As explained in the Introduction, the EAs conducted in the late 1960s and in the 1970s provided some first indications of systematicity in the pattern of L2 acquisition, but this evidence was far from conclusive. More compelling signs of developmental regularities were found in the interlanguage studies conducted from the 1970s onwards, in the shape of morpheme acquisition studies (section 2.1) and of studies of grammatical subsystems such as syntactic structures (e.g. negation, question formation) and the verb system (section 2.2).

As mentioned earlier (section 1), it has often been assumed that the results from such studies apply not just to the acquisition of productive L2 competences but to language acquisition in general – including reception, as this was, and mostly still is, viewed as preceding and being necessary for productive L2 grammar acquisition. To conclude this section on empirical evidence, however a pioneering study investigating systematicity in the development of receptive interlanguage systems is discussed (section 2.3).

Many scholars studying developmental patterns have taken a research-then-theory approach (Goldschneider & DeKeyser, 2001). This literature review too will reserve the explanation of developmental patterns for a later section.

2.1 Early Evidence for Developmental Patterns: Morpheme Studies

2.1.1 Empirical Evidence

In the 1960s Chomsky's mentalist approach to L1 acquisition gave rise to the *independent grammar assumption*, which premises that children are not 'defective speakers of adult language' (Cook, 1993, p. 14; Goldschneider & DeKeyser, 2001, p. 5) but rather speakers of their own distinct languages. In response to this *independent grammar assumption*, various researchers set out to describe these child grammars (Cook, 1993; Goldschneider & DeKeyser, 2001). One such scholar was Brown, whose 1973 study of the acquisition of 14 grammatical morphemes in the spontaneous L1 speech of 3 American children revealed that the morphemes in question are acquired in a more or less invariant order (Brown, 1973; Cook, 1993; Goldschneider & DeKeyser, 2001).

Inspired by Brown's study and by other L1 morpheme studies (e.g. de Villiers & de Villiers, 1973), scholars in the field of SLA began replicating these morpheme studies with the aim of determining whether morpheme acquisition orders also existed for L2 acquisition, and if so whether this order was similar to the L1 morpheme order.

Dulay and Burt (1974), for example, examined the acquisition of 8 grammatical morphemes, using the Bilingual Syntax Measure (BSM) to elicit L2 speech, and concluded that in child L2 acquisition, grammatical morphemes are acquired in 'approximately the same order' by every child (p. 52), irrespective of the L1 background. This order however did not correspond with the morpheme order found in L1 morpheme studies. Similarly, Bailey, Madden and Krashen (1974) found that adults (of 11 different L1 backgrounds) acquired L2 morphemes in language production in an order that correlated with the order found by Dulay and Burt (1974). Some errors which were due to the L1 were observed, but overall the L1 was considered less significant than the 'universal language processing strategies' (ibid., p. 242) which were claimed to be at work. These findings were further corroborated by Pica (1983), who broadened the focus from naturalistic L2 learners to instructed L2 learners. Thus, not only L1 but also

learning context was taken to have no *significant* influence on the acquisition of English morphemes.

As a last example, Larsen-Freeman (1975), who was critical of the fact that many studies used the BSM, extended her own study on morpheme acquisition to reading, writing, listening, imitating and speaking (Larsen-Freeman, 1975). She concluded that the results for the oral production tasks correlated highly with those found by Dulay & Burt. The results for the written production tasks and for the comprehension tasks however did not (ibid.).

Table 1 shows the acquisition order of English morphemes in L2 production, as summarised in recent overviews (e.g. Ellis, 2008; Ortega, 2009a).

Table 1 *Acquisition Order of English Morphemes in L2 Production*

Acquisition Rank	Morpheme
1	-ing
	Plural -s
	Be copula
2	Be Auxiliary
	Article
3	Irregular past
4	Regular Past
	3rd Person Singular
	Possessive -s

Ortega (2009b) acknowledges that within each of these 4 rank groups the order of acquisition may vary across studies, but claims that no variation has been reported across these four acquisition ranks. Apart from this minor variation, the accuracy order 'has been shown to be similar for both young and adult L2 learners, for both naturalistic and instructed learners, regardless of L1 background or whether the data are oral or written' (2009b, p. 85).

2.1.2 Criticisms of the Design and Methodology of Morpheme Studies

The earliest morpheme studies involved Suppliance in Obligatory Context (henceforth SOC) (see 1.1.1), a type of Performance Analysis whereby a particular morpheme is said to have been acquired when it is used in 90 (or sometimes 80) percent of the situations where its use is obligatory (Long & Sato, 1984). It has been argued that by only looking at correct use in obligatory contexts and neglecting incorrect use in other

contexts, SOC produces an incomplete picture of the IL system (ibid.). In later morpheme studies, for that reason, SOC was replaced by target-like use analysis (e.g. Pica, 1983), a method which does take inappropriate use outside of the obligatory contexts into account. Nevertheless, several studies have shown that target-like use analysis and SOC yield the same morpheme acquisition orders (Ellis, 2008).

Another criticism directed at the morpheme studies concerns the cross-sectional design of the majority of morpheme studies. Although many cross-sectional studies make claims about a morpheme *acquisition* order on the basis that 'the more accurately a morpheme was used, the earlier it must have been acquired' (Ellis, 2008, p. 82), their results merely present an *accuracy* order (ibid.). This criticism is supported by certain longitudinal studies (e.g. Rosansky, 1976; Hakuta, 1974; see also Ellis, 2008 for an overview) in which the attested morpheme acquisition orders diverge from the accuracy orders found in cross-sectional studies.

Scholars have also remarked that the evidence for developmental patterns has come from a limited number of morphemes, the results of which have been generalised, and that morpheme studies obscure individual variation by grouping results and calculating mean scores (Cook, 1993; Gass & Selinker, 2008). Further, morpheme studies have rarely looked at L2s other than English so that claims about a 'universal' acquisition order must be treated with caution (Ellis, 2008).

Finally, it has been noted that many of the morpheme studies dismissed the influence of L1 transfer. According to Ellis (2008), however, the rejection of behaviourism and CAH in favour of mentalist views led to a minimalist view on transfer, so that some scholars were 'too ready to reject transfer' (p. 363) and based their rejection on 'flimsy evidence' (ibid.). In later years, the dichotomous view on transfer as either existing or not was abandoned. Ellis (2008) writes that 'transfer is now treated as a mental process in its own right' (p. 54), which 'works alongside other universal and developmental factors' (p. 371).

In spite of the above criticism and in spite of some of the discrepancies in the results of morpheme studies, scholars still argue that 'the morpheme studies have been and continue to be influential in our understanding of the nature of developmental sequences' (Gass & Selinker, 2008, p. 135).

2.2 Evidence from the Study of Grammatical Subsystems

2.2.1 Empirical Evidence

The majority of L2 morpheme studies were conducted in the 1970s and 80s. From the 70s onwards, scholars have also investigated the acquisition of single grammatical subsystems. In particular, the expression of negation, the acquisition of relative clauses, of question formation, of the verb system (i.e. tense and aspect) and of word order in general have been extensively studied (Ellis, 2008).

The aim of this study is not to discuss the development of each aspect of grammar in great detail. Rather, as an illustration of the developmental stages found in productive L2 grammar acquisition, two grammatical subsystems only will be singled out for attention, notably the acquisition of the verb system and of word order in general. The selection of these two examples is not altogether random. Rather, explanatory theories have been developed to account for the developmental patterns in both subsystems, so that both systems will return in section 3.2.2.

In research on the verb system, first, two types of analysis can be discerned, namely meaning-based analyses and form-oriented studies (Bardova-Harlig, 1999; Ellis, 2008; Ortega, 2009a; 2009b). Meaning-based (or concept-oriented) analyses focus on how learners' expression of temporality evolves over time. Research has shown that in a first phase learners use only pragmatic devices to express the temporality – e.g. by using chronological word order. In a second phase learners also make use of lexical devices, such as connectives, to express different types of temporal meaning. It is not until the third and final phase that the verb is morphologically marked (Bardova-Harlig, 1999; Ellis 2008; Ortega, 2009a; 2009b).

Form-oriented studies, on the other hand, have brought forward evidence both for a developmental order in the acquisition of the various verbal morphemes and for developmental sequences in the acquisition of individual morphemes. With regard to the order of acquisition, Ellis (2008) notes that Germanic and Romance language show largely corresponding developmental patterns. For example, one study represents the order of acquisition for English verb morphology as '*past > past progressive > present perfect > past perfect'* (Bardovi-Harlig, 2000 as cited in Ellis, 2008, p. 90) while another study reports the French verb morphology to be acquired in the order '*passé*

compose > '*veux'/'va'* + *infinitive* to refer to the future > *imparfait* > *past perfect'* (Schlyter, 1990 as cited in Ellis, 2008, p. 90). Klein's (1995) order of acquisition also shows some correspondence with the above, namely *third person –s* and *present tense copula* > *irregular past* and *–ing form* > *present perfect* > *regular past* > *future with 'shall' or 'will'* > *past perfect* (see also Ellis, 2008).

With regard to the sequences of development involved in the acquisition of the separate tense-aspect morphemes, finally, it has been observed that once learners have acquired a new verbal morpheme, the use of this morpheme is often overgeneralised – even to linguistic contexts were an earlier-acquired morpheme is required – before it is used only in the appropriate contexts (Ellis, 2008).

Secondly, a large body of evidence for developmental patterns in the acquisition of syntax comes from the ZISA project (Zweitspracherwerb Italienischer und Spanisher Arbeiter), which was reported on in several studies by Meisel, Clahsen, and Pienemann in the early 1980s. The studies conducted in the framework of ZISA examined Italian and Spanish learners of German living in Germany (Ellis, 2008; Pienemann, 1998). On the basis of the ZISA research, Meisel, Clahsen, and Pienemann came up with a clear developmental pattern for the acquisition of German syntax, as can be seen in Table 2. Note that Table 1 is based on the emergence of syntactic structures rather than on the accuracy of usage (as was the preferred method in the morpheme studies).

Table 2 *Word Order Acquisition in L2 German*
(adapted from Ellis, 2008; Ortega, 2009a; Pienemann, 1998)

Acquisition Rank	Name	Explanation
1	Canonical Word Order	Subject-Verb-X
2	Adverb Preposing	X-Subject-Verb (adverb in initial position without the required Subject-Verb inversion)
3	Verb Separation	Verb…Auxiliary/Complement or Particle (non-finite verbal elements are moved into clause-final position)
4	Inversion	X-Verb-Subject
5	Verb Final	Final Position for Verbs in Subordinate Clauses

Slight variations are possible with regard to the speed with which learners progress through the different stages. In addition, the Canonical Word Order with which learners start out typically reflects the L1, so that differences among learners of different L1 background will be found. Finally, not all learners reach the final stage(s) (Ellis, 2008).

As said, the acquisition of other grammatical systems will not be discussed here. In short, the analysis of different grammatical subsystems – both in English and in some cases in other languages as well – has brought forward abundant evidence for the existence of orders and sequences of development in language production. With regard to the role of the L1, it is noteworthy that whereas scholars used to take a dichotomous view on transfer, this is now no longer the case. As mentioned above, many studies of grammatical subsystems observe certain L1 effects. These effects include for example avoidance, differential learning rates, differences in the path of acquisition and overproduction (Gass & Selinker, 2008).

2.2.2 Design and Methodology of the Studies

In the studies of grammatical subsystems, the Performance Analysis method which was used in the morpheme studies but which was subject to stringent criticism is replaced by new methods. A first method is Frequency Analysis or Interlanguage Analysis (Ellis, 2008). As Ellis (2008) explains, Frequency Analysis 'involves identifying the variants of a given structure and examining the frequency of occurrence of each variant' (p. 963). It is used for analysing developmental *sequences* (see Terminology). This method constitutes a considerable improvement compared to EA in that it does not reflect the Comparative Fallacy; i.e. it does not analyse IL in relation to native-speaker language (see 1.1.2).

Scholars also frequently made use of Implicational scaling when conducting cross-sectional studies (Ellis, 2008). As Ellis (ibid.) explains, implicational scaling 'exploit[s] the inter-learner variability (…) in order to establish which features different learners have acquired and whether the features can be arranged into a hierarchy according to whether the acquisition of one feature implies the acquisition of (…) other features' (p. 69).

As final tokens of improvement, Ellis (2008) observes that the studies discussed in this section cannot be criticised for overgeneralization since they study a grammatical subsystem and only apply their findings to this one subsystem; and further that, in contrast to morpheme studies, which only looked at the acquisition of English morphemes, studies of grammatical subsystems have broadened their interest to other languages, thus presenting interesting findings concerning the *universal* validity of developmental patterns.

2.3 Interpretation of the term 'Developmental Patterns'

As Ellis (2008) points out, the notion of developmental patterns can be interpreted with several nuances. First, the term can refer either to the development of specific L2s only or to universal patterns valid for L2s. Morpheme studies provided evidence for the former type of developmental patterns, as only English morphemes were examined. Evidence for the latter is scarcer but can be attested in the acquisition of verbal morphology.

Secondly, Ellis (2008) also notes that developmental orders and sequences are not necessarily 'rigid'. As has also been mentioned above, some variation within the developmental patterns is possible, particularly as a result of the learners' L1 background or as a result of individual learner variability. This variation concerns for example the speed with which learners progress through the different developmental stages, but may also be reflected in the addition or by contrast the missing out of certain developmental stages.

Finally, it warrants emphasising that developmental patterns – at least in a more rigid interpretation – are not yet universally accepted. Lightbown (1984, as reported by Ellis, 2008) claimed that 'for every study that reports an order or sequence, there is another study which has produced counter-evidence' (Ellis, 2008, p. 112-113). According to Lightbown, furthermore, there may well be 'universal tendencies' (Ellis, 2008, p. 113), but the influence of the learners' L1, of L2 input, and of the L2 learners' social-psychological attitudes should not be underestimated.

2.4 Developmental Patterns in Receptive Language Acquisition?

It has been amply stated in the previous sections that language reception has been largely overlooked in investigations concerned with the development of L2 grammar. In the morpheme studies, for example, scholars looked at data gathered through production tasks (section 2.1). One exception was Larsen-Freeman (1975), who, apart from oral and written production task, also looked at the participants' performance in reading and listening tests. Although there did appear to be a high degree of concordance across participants in the receptive tasks, the morpheme orders in these tasks did not correlate with the morpheme orders found through productive language tasks.

The discussion of the relationship between the acquisition of receptive and productive competences (Chapter I, Section 1) further confirms that it would be premature to assume that the acquisition of receptive language skills, like the acquisition of productive skills, will contain developmental patterns, or if these developmental patterns exist, that they will be similar to the ones found in productive L2 grammar acquisition. Many researchers have (tacitly) presumed, and for the most part still presume, that this is the case. However there is also evidence that the relationship between the acquisition of language reception and language production may be more intricate than has hitherto been assumed.

The lack of attention for the development of receptive L2 grammar knowledge has recently been addressed by Berthele, Kaiser and Peyer (in press). In their study, Berthele et al. seek to find a (tentative) answer to whether there are 'systematic and/or continuous developments (…) regarding the comprehension of items containing particular grammatical structures' (in press). In view of the difficulties involved in examining the hidden corpus of receptive grammar knowledge, the study also constitutes an attempt to determine 'if it is methodologically possible to empirically investigate this question' (ibid.).

The participants involved in Berthele et al.'s study were Italian- or French-speaking adults with a high proficiency in English (L2) and who were learning German as a 3rd of nth (n>2) language. The study focused on the comprehension of a number of German grammatical categories in written texts, taking into account factors such as L1 transfer and paying attention to the processing strategies used (for example, semantic processing).

More specifically, participants were given a text, written in the style of an encyclopaedia article, in which potentially problematic grammatical categories were included. The aim of the text was to find out whether the presence or absence of these grammatical categories had an effect on the comprehension of the text. To control for the lexical factor, a translation of the content words was available to the participants.

An examination of the error rate of the comprehension of the grammatical categories showed a variegated picture, with differences among the categories in the degree of progress. Roughly speaking Berthele et al. were able to distinguish three stages, namely

a beginning stage, a stage of instability and restructuring (characterised by U-shaped behaviour) and a stage of gradual progress.

In addition to examining error rates, Berthele et al. looked the effect size of certain grammatical categories on text comprehension at different proficiency levels. The picture this presented was very complex, and according to Berthele et al. did not yet allow definite conclusions with respect to whether or not there is systematicity in receptive grammar development. Even so Berthele et al. ventured to say that 'some of the regularities discovered in the development of learners' production data might indeed apply to the receptive channel as well' (in press).

2.5 Conclusion

To recapitulate, it has been shown that evidence for the existence of developmental patterns in productive L2 acquisition abounds. The morpheme studies conducted in the 1970s contained the first clear sign of systematicity in productive L2 grammar acquisition, but their validity is now questioned. More recent and less disputed evidence has been found in studies of grammatical subsystems, such as the verb system or syntactic aspect of grammar. These studies suggest that developmental patterns exist, in the shape of relatively fixed acquisition orders – referring to the acquisition of various L2 grammar features one after another – and sequences – referring to the stages (or *sequences*) in the acquisition of particular grammar features.

Developmental patterns have not yet been universally accepted– though scholarly agreement is widespread. Where developmental patterns are accepted, furthermore, opinions vary with regard to the universal applicability of particular patterns. Finally, it has been noted that developmental patterns need not be rigid. Some individual variation need not invalidate the evidence.

Throughout the studies discussed above it has been (tacitly or explicitly) assumed that developmental patterns apply to both receptive and productive grammar development. Empirical evidence for this is non-existent however. One study has recently examined the comprehension of grammatical features in written text, but although it was suggested that there may be some similarity between the acquisition of L2 grammar knowledge in language reception and language production, the evidence was not

detailed or conclusive enough to allow far-reaching conclusions. Clearly receptive L2 grammar acquisition is a topic worthy of future investigation.

3 Explaining Developmental Patterns

3.0 Introduction

Section 2 has given an overview of the evidence supporting the existence of developmental patterns in the acquisition of L2 grammar. Notwithstanding the objections raised against some of these studies – in particular the morpheme studies – and the concession that it is not always tenable to speak of rigid and universal developmental patterns, the notion of developmental patterns has found wide acceptance in SLA.

Researchers investigating these regularities in the development of L2 grammar have long taken a research-then-theory approach. This was most striking in the morpheme studies, where explanations were frequently limited to appeals to 'universal language processing strategies' (Dulay & Burt, 1974, p. 52). Theory-driven research began to occur much more recently (Ellis, 2008).

In this section, some of the explanations that have been offered to account for the systematicity attested in L2 grammar acquisition will be discussed. A first subsection contains some theories and explanations which have been developed with the specific aim of accounting for (specific) development patterns and hence belong to the research-then-theory practice mentioned above.

According to Ellis (2008) 'any theory of L2 acquisition will need to account for developmental patterns' (p. 113). In a second section therefore some theories of language acquisition will be presented which, as part of their much wider-scoped domain of interest, have provided an account for developmental patterns.

As a final note, it may be noted that it is neither possible nor necessary within the scope of this study to give an exhaustive overview of the explanations that have been offered for developmental patterns. For the same reason, the explanations that are included in this overview cannot be discussed in great detail. Therefore, for more in-depth discussions of the theories and hypotheses offered in this overview and form a discussion of explanatory accounts that have not been included, the reader is referred to

wider-scoped overview works or to monographs on the relevant topics (such as Ellis, 2008).

3.1 Research-Then-Theory: Explanatory Theories for Developmental Patterns

3.1.1 Frequency Effects (and Other Grammatical-Item Inherent Factors)

The morpheme studies – the validity of which is now subject to some debate but which are nonetheless worth discussing in this section – will be taken as the starting point of this section. As explained in the introduction, many morpheme studies of L2 acquisition were predominantly descriptive, giving very little attention to possible explanations of their findings. Arguably, it is owing to the shift from behaviourist to mentalist views on L2 acquisition that in many morpheme studies the newly-developed notions of innate language capacities and 'universal language processing strategies' (Baily, Madden, and Krashen, 1974, p. 242; Dulay and Burt, 1974, p. 52) – however vague these terms – were deemed sufficient in explaining systematicity in L2 grammar acquisition. Dulay and Burt (1974) even look upon the vagueness of these universal language processing strategies as a necessity:

> 'It would be very tempting to formulate strategies based on the acquisition sequence we obtained. However, we believe that "universal strategies" should be sufficiently abstract and comprehensive so as to predict acquisition orders based on different types of languages, such as languages other than English, or types of speech exposure other than natural speech.' (p. 52)

In the field of SLA, Larsen-Freeman (1976) is named as the first one to attempt to find an explanation for the morpheme study results (e.g. Goldschneider & DeKeyser, 2001; Ortega, 2009a). By examining a variety of factors such as syntactic and phonological complexity, affective variables, operating principles, perceptual saliency, and frequency, Larsen-Freeman ultimately came to the conclusion that frequency of occurrence is the only plausible predictor for the L2 morpheme acquisition order. However, she wrote that her conclusion 'must remain tentative until second language acquisition orders are compared with morpheme frequency counts from the speech of a larger sample of native speakers using different registers in conversing about a variety of topics' (p. 132).

Larsen-Freeman thus provided tentative support for the Frequency Hypothesis posited by Hatch and Wagner Gough (Larsen-Freeman, 1974), which – as its name suggests – holds that the order of acquisition of linguistic items 'is determined by the frequency with which [they] occur in the input' (Ellis, 2008, p. 242). Numerous empirical studies have corroborated the influence of frequency, not just for morpheme acquisition orders but also for other aspects of grammar acquisition and of L2 acquisition more generally (e.g. vocabulary) (for an overview: see Ellis, 2008).

While frequency is thus an undeniable factor in L2 grammar question, it has also been amply demonstrated that it is not the only factor. Goldschneider and DeKeyser (2001), in what is considered the first comprehensive and plausible attempt to explain morpheme acquisition orders (Ortega, 2009b), conducted a 'meta-analysis' of 12 morpheme studies which identified 5 factors which, together, 'account for a large percentage of the variance in order of acquisition' (p. 36), i.e. perceptual salience, morphophonological regularity, semantic complexity, syntactic category and frequency of occurrence. Alternatively, these five factors can be seen as 'phonological, morphological, syntactic, semantic and numerical aspects of salience' (p. 36). According to Goldschneider and DeKeyser (2001) their research shows that other explanations – such as Pienemann's Processability Theory (discussed below in section 3.1) or 'innate blueprints' as in Universal Grammar (section 3.1.4) - are 'not required to explain order of acquisition' (p. 38).

Objections can be raised against Goldschneider and DeKeyser's claim. First, it should be noted that the empirical studies which they examined in their meta-analysis only pertain to morpheme orders and hence do not necessarily explain the acquisition of syntax. Secondly, as Goldschneider and DeKeyser (2001) concede, their results only apply to English as a second language. Finally, Goldschneider and DeKeyser (2001) themselves admit that 'the combination of factors used in this study does not account for all the variance in accuracy scores' (p. 38).

Finally, frequency effects have also been investigated by N. Ellis (2002). In a review of research pertaining to 'phonology, phonotactics, reading, spelling, lexis, morphosyntax, formulaic language, language comprehension, grammaticality, sentence production, and syntax' (p. 143) N. Ellis – who is an adherent of Connectionism, as will be discussed in section 3.2.2– concludes that frequency is an 'all-pervasive causal factor' (p. 179) in L2

acquisition. Nonetheless, N. Ellis also admits that frequency alone does not suffice as a determinant factor. Hence he agrees with Goldschneider and DeKeyser (2001) that a multitude of factors interact with frequency in the L2 acquisition process.

Finally, it is worth noting that neither N. Ellis (2002) nor Goldschneider and DeKeyser (2002) make an explicit distinction between language production and language reception. Hence it cannot be said whether both aspects of language acquisition are implied in their discussion of L2 acquisition. Larsen-Freeman (2002), by contrast, in a response to N. Ellis' (2002) article, writes – albeit as a side note only – that we should '(…) not assume an isomorphism between production and comprehension grammars' (p. 283).

3.1.2 Markedness

In SLA the concept of markedness has a different meaning depending on whether it is used in the context of Universal Grammar or of typological language universals. In this section, only the concept of markedness as underlying language typology will be discussed. The Universal Grammar related meaning of markedness will be discussed in section 3.2.1

A general definition of markedness says that 'linguistic phenomena that are common in the world's languages, that seem easier for linguistic processing, and that are more "natural" than other, are unmarked as opposed to their marked variants' (Hyltenstam, 1986, p. 57). In language typology, markedness is seen as a relative concept, with certain linguistic phenomena being more or less marked than others (Ellis, 2008, p. 561).

Markedness has been investigated as a potential determinant of acquisition orders in the development of various grammatical items, such as the position of the negator, preposition stranding (i.e. the spatial separation of a preposition and (pro)noun) and pied piping (i.e. the opposite of the former, namely the co-positioning of these two elements). In research focusing on developmental patterns in the acquisition of relativisation, furthermore, 'the most widely investigated influence stems from the typological concept of markedness' (Ortega, 2009a, p. 91). For example, Hyltenstam (1984, 1986) argues in favour of markedness as an influential factor in the acquisition of relativisation.

More specifically, the role of markedness in the acquisition of relativisation is described in the Noun Phrase Accessibility Hierarchy (NPAH) developed by Keenan and Comrie. In this theory six universal types of relative clauses are discerned, based on the function of the relative pronoun (e.g. Subject, Direct Object). According to the NPAH, these types of relative clauses 'are in a markedness relationship which is hierarchical and implicational' (Ortega, 2009a, p. 93): both within languages and within learners' interlanguages, 'each lower (more marked) type is possible only if all other preceding (less marked) types are also possible' (ibid).

Whereas Ortega (2009b) argues that scholars have been very successful in applying the markedness concept to the acquisition of relative clauses, Ellis (2008) takes a more sceptical stance. He questions not only the NPAH, but also more generally whether markedness can truly explain the acquisition of relative clauses, or of other grammatical items to which the concept has been applied. First, although the concept of markedness may have been shown to have an influence on both the ease with which certain L2 items are learned and on the order in which they are learned, and as a consequence should not be dismissed, Ellis (2008) emphasis that markedness alone is not a sufficient explanation. Secondly, because 'we can assume that unmarked features are likely to be frequent' (p. 557), there is no certainty as to whether markedness or rather input frequency determines the acquisition of a linguistic item.

3.1.3 Tense and Aspect: Aspect Hypothesis

As mentioned in 2.2, a major aspect of grammar acquisition in which developmental regularities have been attested is the acquisition of tense and aspect. More specifically, three phases have been discerned which are traversed by many L2 learners: a pragmatic phase, a lexical phase and a grammatical phase. Within the third phase, moreover, developmental patterns have been attested in the order in which verbal morphemes are acquired.

One hypothesis suggests that the acquisition of verbal morphemes within this third phases is governed by 'the inherent aspect or lexical semantics' of verbs (Ortega, 2009a, p. 86). The Aspect Hypothesis is based on Vendler's four-way classification of verbs, on the basis of their semantics, into verbs expressing 1 achievement (i.e. a punctual event) 2 accomplishment (i.e. a telic event, having duration but with an inherent endpoint, such as *run a mile*), 3 activity (i.e. having duration but no *inherent* endpoint,

such as *run*) and 4 state (continuing but having no dynamics and requiring no effort or energy, such as *see* or *love*) (Andersen & Shirai, 1996). For example, the –ing form will initially be used with verbs expressing duration only, while past and perfect marking will be used only with verbs expressing accomplishment or achievement (Ortega, 2009a, 2009b; Andersen & Shirai, 1996).

Empirical studies have been conducted to test the Aspect Hypothesis (e.g. Robison, 1995; Rohde, 1996), with varying results. For example, Rohde's (1996) study of naturalistic L2 learners found that the Aspect Hypothesis is invalidated at least for some verb inflections. More particularly, 'the distribution of verbs occurring in the progressive form does not comply with the hypothesis' (p. 1133). Nonetheless, the Aspect Hypothesis is usually regarded as an acceptable explanation for the acquisition of tense and aspect, albeit in combination with other factors such as L1 transfer (e.g. Ellis, 2008; Ortega, 2009a; 2009b).

The Aspect Hypothesis will not be discussed in more detail here because the scope of this hypothesis is limited to the acquisition of the verb system. Notwithstanding this limited applicability, it has been worth discussing this hypothesis here because it contributes to illustrating the diverging natures of the hypotheses developed to account for L2 grammar acquisition patterns. Moreover, the Aspect Hypothesis receives support from a cognitive theory of L2 acquisition called Operating Principles which will be discussed later (this chapter, section 3.2.2).

3.1.4 Processability Theory

Pienemann's Processability Theory (henceforth PT) (1998; 2005a; 2005b) is a cognitive account of developmental patterns in L2 grammar acquisition. It emerged out of the Multidimensional Model, which was developed to explain the empirical findings (i.e. developmental patterns in the acquisition of German syntax) of the ZISA project (Ellis, 2008; Pienemann, 1998; see section 2.2). As Ortega explains, in cognitive theories of language 'the human mind is viewed as a symbolic processor that constantly engages in mental processes' (p. 83). PT then is a cognitive theory of language acquisition which seeks to explain the developmental patterns in language acquisition by appealing to 'the architecture of human language processing' (Pienemann, 2005a, p. 2). PT will be discussed in some detail because it is considered one of the most comprehensive accounts of developmental patterns.

The 'point of departure' for PT is Levelt's psycholinguistic Model of Speech Production (Pienemann, 1998, p. 54) which distinguishes three stages in language production. In the *conceptualizer* the language speakers conceptualises his/her message (Levelt, 1989; as cited in Pienemann, 1998, p. 54). In the *formulator*, the 'preverbal message' created by the conceptualizer is 'translate[d] (…) into a linguistic structure' (ibid.). The *articulator* converts the formulator's output into speech. Pienemann's theory is concerned only with what happens in the formulator.

On the basis of Levelt's theory, Pienemann posited the following basic premises. In creating the speech act, a number of *processing components* are at work. These processing components work autonomously, automatically and in parallel and 'are able to accept and pass on only information of a highly specific nature, for instance, only concerning noun phrases' (Pienemann, 2005a, p. 5). In this construction process, grammatical information has to be stored in memory. For example, in processing the sentence *She gives him a book* information concerning the person and number of the subject has to be stored in order to be able to provide the verbal agreement marker (Pienemann, 2005a). Because Levelt's formulator is said to work automatically and because production involves the processing of large amounts of information at high speed, it is argued that working memory cannot carry out the memory storage task. Instead, it is posited that a task-specific grammatical memory store (the Syntactic Buffer) exist (Pienemann, 2008, p. 60).

As Ortega (2009b) explains, PT's explanation of developmental patterns hinges on the learner's capacity to store grammatical information 'within and across units in the linguistic material they encounter'. According to PT, developmental patterns are, 'at least in part' (Pienemann, 2005a, p. 3), the result of cognitive processing constraints which prevent the learner from producing certain language structures. That is to say, it is posited that a learner cannot produce a particular language structure unless s/he has acquired the processing procedure (or computational routine) required to produce the linguistic structure. Acquiring a L2 thus 'includes the acquisition of the procedural skills needed for the processing of the language' (Pienemann, 2005a, p. 2).

Pienemann (2005a) distinguishes 5 language-specific processing procedures which have to be acquired in order for native-like utterances to be possible. The processing procedures are ordered in an implicational hierarchy and consist of (1) lemma, (2)

category procedure, (3) phrasal procedure, (4) S(entence)-procedure and word order rules and (5) matrix/subordinate clause. Although *lemma* is included in this hierarchy, this phase refers not to an actual processing procedure but rather to the absence of any processing procedures.

The distinction between these 5 processing procedures has to do with the type of grammatical information that can be stored and exchanged. In a first phase, called *lemma* in the processing hierarchy, the learner is unable to store any grammatical information, so that output is restricted to single word utterances which are invariant in form (e.g. no plural marking is possible). In a second phase, the learner acquires the *category procedure*, enabling him/her to mark the words for example for number or tense. In a third phase, the *phrasal procedure* allows the learner to exchange information (by storing this information in memory) between the different constituents of a phrase. In the *S-procedure*, exchange of information within a sentence, i.e. between the heads of different phrases, is possible. Finally, the procedure *matrix/subclause* allows the learner to process subclauses. In the first three phases, technically, sentences cannot be formed, but learners can for example produce unanalysed chunks. Also when the learner has access to the *category procedure* but not to the succeeding categories, sentences with a canonical word order (SVO) can be formed.

To conclude, PT is one of the most comprehensive explanations for L2 acquisition patterns, for a number of reasons. First, it accounts for developmental patterns in the acquisition of both morphology and syntax, in both the L1 and the L2. Pienemann admits that developmental patterns in the L1 and L2 have been shown to be different, but argues that these dissimilarities do not mean that the L1 and L2 are processed in different ways. Rather, the initial hypotheses of L1 and L2 learners are different (for example, canonical word order may be SVO or SOV) and this affects the subsequent acquisition of the category procedures (Ellis, 2008). Further, although the particularities cannot be discussed within the scope of this study, it may be remarked that PT also accounts for learners' deviation from the developmental patterns and for differences with regard to rate of acquisition and ultimate attainment. Finally, PT also addresses the issue of L1 transfer, claiming that L1 transfer only occurs when the transferred grammatical item can be processed by means of the processing procedures already acquired by the L2 learner (Ellis, 2008).

PT was originally developed to account for the developmental patterns in the acquisition of German and English syntax and morphology among learners with German L1s, but Pienemann nonetheless claims that PT has universal value (Pienemann, 1998, 2005a). No empirical evidence for this was provided by Pienemann himself in his original study, however, with the exception of a discussion of two studies on the acquisition of Japanese as a L2. As neither of these studies was conducted with the specific aim of testing the PT, they only provide preliminary evidence. Pienemann himself indeed observed that 'this is merely a first look at the acquisition of Japanese' (p. 207). However several studies aimed at assessing the validity of this universality claim have shown that PT may indeed apply to languages other than German and English – both typologically similar languages such as Swedish (e.g. Pienemann and Håkansson, 1999) and typologically distinct languages such as Japanese or Italian (e.g. Di Biase and Kawaguchi, 2002; Sakai, 2008).

One respect in which PT falls short of comprehensiveness is in the lack of distinction between language production and language reception. As Ellis (2008) rightly remarks, PT 'is in actuality a theory of language production' (p. 461) because it is based on Levelt's model of speech production. Pienemann himself writes that PT accounts for 'describable developmental routes' (1998, p. 4). Since the supporting data for these 'describable' routes only comes from language production data (see section 2 above), and since Pienemann himself presents no empirical evidence with regard to language reception, it is wise to take a cautious stance and not automatically presume that PT also applies to the developmental of receptive language skills. In this respect, however, PT is not any different from other theories or hypotheses which seek to explain PT.

3.2 Theories of Second Language Acquisition

The previous section contained a number of theories developed with the explicit intent of accounting for (specific) developmental patterns in productive L2 acquisition.

Explanations for developmental patterns can also found in some L2 acquisition theories. Ellis (2008) broadly distinguished three types of L2 acquisition theories, namely cognitive theories, linguistic accounts and sociocultural theory. The latter, claims that 'learning originates in social interaction', and as such 'constitutes a challenge to the claimed universality of the order and sequence of acquisition' (Ellis, 2008, p. 549).

Although any theory of L2 acquisition should account for the route of L2 grammar acquisition, thus far only a few of the many cognitive theories that exist, give a (partial) explanation for the existence of developmental patterns. These will be discussed in section 3.2.2. First, however, the account offered by Universal Grammar, a linguistic theory, will be discussed.

3.2.1 A Linguistic Account: Universal Grammar

Very briefly, the notion of Universal Grammar (UG) was developed by Chomsky to account for the 'logical problem of language acquisition' or the problem of learnability, namely the observation that the 'crumbs and snippets of speech' (Aitchison, 1998, p. 103) which form children's L1 input cannot account for their successful acquisition of the L1 and their capacity for creative language usage, in particular given their 'lack of cognitive sophistication' (Ellis, 2008, p. 592). Chomsky inferred from this that children must have an innate knowledge of the L1 grammar which constrains the form a language construction can take (Ellis, 2008; Ortega, 2009b). It is this knowledge the term UG refers to.

UG is described in terms of *principles*, i.e. 'highly abstract properties of grammar which govern all languages' (Ellis, 2008, p. 982). Some of these principles are *parameterized* - meaning that they have different settings. These parameters account for the differences among languages. On the basis of the evidence provided by input, the principles are activated (or 'triggered') in the child's mind and set in the correct parameter. An example of a principle is that all language structures have a head and modifier (Aitchison, 1998). Input is needed to activate the knowledge of this principle and to determine the position of these modifiers.

Whether or not the 'logical of problem of language acquisition' is also applicable to L2 acquisition is the matter of some debate, since L2 learners in many cases do not reach native-like competence. Consequently, there is also no scholarly agreement on the role played by UG in L2 acquisition – with opinions taking either a 'full access', 'partial access' or 'no access' stance.

More interesting for the sake of this study is 'the developmental problem of L2 acquisition', which is concerned with determining the role of UG in the developmental

patterns attested in the acquisition of L2 grammar. There are a number of possible ways in which UG is theorized to influence the orders and sequences of L2 acquisition.

A first theory, called the maturational view or the Maturation Hypothesis, argues that different parameters and principles become accessible at different moments in time (Clahsen, 1992; Ellis, 2008; Felix, 1992). According to Clahsen (1992) the Maturation Hypothesis comes in a strong and a weak version. The strong version maintains that the order in which the principles and parameters becomes accessible is determined by an innate schedule. The weak version appeals to 'UG-external constraints' which restrict the availability of principles, up to the moment that these constraints are, as a result of maturation, 'lost' (p. 54).

Another commonly found view with regard to the role of UG in determining the route of L2 grammar acquisition is that UG 'interacts with other cognitive mechanisms' (Ellis, 2008, p. 596). In this view it is argued that L2 input can trigger a UG principle only when the language parser mechanism, which processes input, is capable of 'perceiving the relevant input data' (p. 579).

In yet another perspective it is assumed that UG is in its entirety accessible to L2 learners from the initial stage of language development. However the order in which UG principles are acquired is affected by the markedness of the different parameter settings. Note that the meaning of the term *markedness* within UG is different from the notion of markedness in research concerned with typological language universals (as described in section 3.1.2). In UG markedness has to do with the degree of evidence needed for acquisition. For example, a distinction is made between core rules, which are governed by principles, and peripheral rules, which exist of 'borrowing, historical residues, inventions, and so on' (Chomsky, 1980, as cited in Hyltenstam, 1986, p. 60). The latter are marked because they 'require (…) substantial evidence' for acquisition (Ellis, 2008, p. 591). Within the core some parameter settings are (more) marked. Unmarked parameters are acquired before marked ones.

Thus, Universal Grammar offers a number of theoretical accounts of why L2 learners pass through regular, universal stages en route to native-like proficiency. The problem with these explanations is that they are difficult to either falsify or validate by means of empirical evidence –as is the case with UG in general (Ellis, 2008).

3.2.2 Cognitive Accounts of Second Language Acquisition

A comprehensive theory of L2 acquisition, which accounts for all aspects of L2 acquisition, does not exist yet, Ellis (2008) writes. In particular, the phenomenon of systematicity in the L2 (grammar) acquisition route is incorporated in only a few theories. Based on the overviews of major, influential cognitive accounts in Ellis (2008) and Gass and Selinker (2008), two cognitive accounts have been singled out for attention – the Connectionist Model and Operating Principles. This does not mean that there are no other theories which account for developmental patterns. However in the above overviews no other theories offering an explanation were encountered.

The Connectionist Model is one of the most influential types of emergentist theories and holds a usage- / exemplar-based view on language learning. The Connectionist Model claims that, on the basis of experience, learners form 'associations between co-occurring elements of the language' (N. Ellis, 2002, p. 173: see also N. Ellis, 2005). In the process of language learning, the strength of these connections is modified in accordance with the evidence provided through input. According to certain models of Connectionism, furthermore, 'rule-like representations may arise out of the network of associations that learners build (R. Ellis, 2008, p. 467). Thus, the Connectionist Model stands in sharp contrast to innatist, rule-based accounts of L2 learning, such as Universal Grammar (Ellis, 2008; Gass & Selinker, 2008).

Connectionism is worth discussing here because it provides support for the role of frequency in L2 grammar learning. Since language learning involves the formation of associations on the basis of input, it stands to reason that the frequency with which linguistic items (co-)occur will influence the strength of these connections. As mentioned in section 3.1.1, N. Ellis, an adherent of Connectionism, has argued in favour of the role of frequency in the morpheme acquisition orders.

The second cognitive L2 acquisition theory discussed in this section is Operating Principles. The term *Operating Principles* refers to learner strategies which are employed, in the course of L1 and L2 acquisition, to process input, so that the information contained in the input can be used to acquire the TL grammar. Operating Principles were first invoked by Slobin (1973) in the context of L1 acquisition and were later applied to L2 acquisition, notably by Andersen (Andersen & Shirai, 1996; see also Ellis, 2008; Ortega, 2009b).

Ellis (2008) argues that Operating Principles offer 'only a limited account of how learners acquire an L2' (p. 456). Furthermore, some criticism has been levelled against the theory of Operating Principles, in particular with regard to how the principles are supposed to be related to each other and with regard to how much weight one is supposed to attach to each of these principles (ibid.). Even so, the Operating Principles are worth mentioning here because they offer support for the Aspect Hypothesis discussed in section 3.1.3.

The Operating Principles supporting the Aspect Hypothesis can be summarized as follows. The *Relevance principle* claims that the acquisition of grammatical morphemes is influenced by the relevance of the morpheme to the meaning of the verb. For example, aspect is more relevant than tense, mood, or agreement. The *Congruency Principle* holds that learners will first use morphemes 'whose aspectual meaning is most congruent with the aspectual meaning of the verb' (Ellis, 2008, p. 457); e.g. progressive markers are used with verbs expressing activity. Thirdly, in accordance with the *One-to-One principle*, 'an intended underlying meaning is expressed with one clear invariant surface form (or construction)' (Andersen, 1984, as cited in Ellis, 2008, p. 457), so that 'the past morpheme is used to signal completion (…), the progressive morpheme ongoing activity' (Ellis, 2008, p. 457). Finally, according to the *Prototype Principle* each morpheme is first used with its most prototypical meaning and 'associate[d] with the most prototypical members of each semantic class of verbs' (ibid.).

3.3 Conclusion

While many - though, as mentioned, not all – scholars agree over the existence of developmental regularities in L2 grammar acquisition, scholarly opinion is much less united over how this systematicity can be explained. In the early years of SLA, scholars were seldom concerned with offering critical explanations for their evidence of developmental patterns. Often a reference to universal mental processes was deemed sufficient. In more recent years, however, scholars have begun to pay more attention to possible explanations – resulting in a considerable number of often contrasting explanations.

Scholars seeking to explain developmental patterns have appealed to the following factors: frequency of occurrence, markedness and cognitive processing procedures.

Frequency of occurrence was first claimed to be an important factor in one of the morpheme studies, but was not considered a highly plausible explanation for developmental patterns until Goldschneider and Dekeyser's meta-analysis of morpheme studies. It is now considered an influential factor, in particular for the acquisition of morphemes, though most likely it works in combination with other factors.

Markedness – in the definition it takes in research dealing with typological universals – has also been claimed to account for developmental patterns, especially in the acquisition of relativisation – as is evidenced by the Noun Phrase Accessibility Hierarchy. Nonetheless, markedness too is probably only one of several grammatical feature-inherent factors exerting influence on the route of L2 grammar acquisition.

An altogether different, but also more comprehensive, explanation is offered by Pienemann's Processability Theory, which maintains that developmental patterns are the result of universal processing constraints. In order to produce language, learners must be able to store grammatical features across increasingly larger sentence constituents. The universal processing procedures required for storing this information, become available in an implicational, hierarchical order, thus causing developmental patterns in grammar acquisition.

In addition to the above theories, which were developed with the explicit aim of explaining specific (e.g. morpheme orders) developmental patterns, an account for the phenomenon of developmental patterns in general is also found in some second language acquisition theories. In theory, all language theories should account for this phenomenon, but to date no such theory exists. Three theories which do offer an explanation for developmental patterns is Universal Grammar theory, Connectionism and Operating Principles. Universal Grammar theorist have offered different types of explanations, all of which are equally difficult to falsify or corroborate. Connectionism offers theoretical support for the role of Frequency. Operating Principles offer a limited account, accounting only for the Aspect Hypothesis, which explains the regularities attested in the acquisition of verbal morphology.

To conclude, like the developmental patterns they seek to explain, the above theories and hypotheses only apply to productive grammar development and can only extrapolated to receptive grammar acquisition if the traditional views on the relationship

between language reception and language production – as explained in section 1 – are legitimate.

4 The Role of Transfer, Learning Setting and Age

This section contains a recapitulation of and brief elaboration on the influences of the L1 – through transfer –, of learning context and of age of acquisition on the L2 grammar acquisition pattern.

With regard to transfer, it is clear from the previous section that the scholarly opinion of the role of this phenomenon has moved from regarding transfer as an all-pervasive factor (i.e. in behaviourism) over a minimalist position (as illustrated by the discarding of transfer in many morpheme studies) to a more nuanced view in which transfer is seen as a complex factor which works alongside other factors in influencing L2 acquisition (Ellis, 2008). Transfer is now also seen in a more complex light than simply as facilitating L2 acquisition where the languages are similar and rendering the process difficult where the languages are difficult. Below is a – far from comprehensive – overview of some of the ways in which L1 transfer is belied to manifest itself.

For example in L2 grammar acquisition research, it is generally accepted that transfer may not just facilitate acquisition or render it difficult. A much more common result is the avoidance of certain grammatical items. Some scholars have also suggested that transfer may affect the rate with which learners progress through developmental stages, and may also have as an effect that certain stages in the developmental pattern are skipped (Ellis, 2008; Gass & Selinker, 2008). It has further been suggested that transfer only occurs once L2 learners have reached a level of proficiency which is high enough for them to be able to notice similarities in the L1 and L2. This may lead to U-shaped behaviour when a learner uses the newly observed similarity to overgeneralize a structure to inappropriate contexts where formerly the universal developmental sequences were followed and the correct L2 structure was used (Berthele et al., in press; Ellis, 2008).

The above discussion merely illustrates the complexity of the notion of transfer, and some major ways in which it may affect L2 grammar acquisition. Important to note is that, despite the general acceptance of transfer as an influential factor in L2 grammar acquisition, the notion of transfer need not contradict the existence of (universal)

developmental patterns. Variation in the pattern of acquisition as a result of individual factors or L1 background does not necessarily invalidate the universality (within or across L2s) of a developmental pattern.

Some scholars have further suggested that the acquisition of L2 grammar proceeds in a universal and systematic manner, independently of the context in which the L2 is learned (for example Pica, 1983; see section 2.1 above). More specifically, this claim centres on the distinction between naturalistic and instructed L2 learners, and on the effect of instruction on the developmental orders and sequences. For an exhaustive discussion, learning context and language instruction, which are broad concepts, would have to be defined in more detail. For example, a distinction would have to be made between meaning-focused instruction, as can be found in immersion education or CLIL, and form-focused instruction, as can be found in formal L2 classes, and between the different subtypes of meaning-focused and form-focused instruction (de Graaff & Housen, 2009). Such a detailed discussion is not possible within the scope of this study, however. Rather this section will have to limit the discussion to the general effect of learning setting of L2 acquisition.

There appears to be a consensus that language instruction can affect the rate of L2 acquisition and the learners' level of ultimate attainment. The route of acquisition, by contrast, is usually posited to be unaffected by the learning setting (de Graaff & Housen 2009; Ellis, 2008). A definite explanation for this cannot be given, since this depends on what causes universal developmental patterns in the first place. If the explanation for developmental patterns is sought in Pienemann's Processability Theory (see section 2.3 below), for example, it could be argued that the order in which processing procedures become available to the learner is impervious to instruction. The rate of acquisition could be affected, in positive or negative sense, by whether the order in which grammatical items are presented to the learner corresponded with the natural order of acquisition or not (Lightbown, 1985; Pienemann, 1985). If one accepts, on the contrary, that developmental patterns are sought in learner-external factors, such as the frequency of occurrence or salience of grammatical items (Goldschneider & DeKeyser, 2001; see also section 2.1.1), an explanation could be found in (possible) differences between naturalistic input and input in instructed learning contexts.

Lastly, an often-debated factor in L2 acquisition which has been suggested not to affect the route of L2 grammar acquisition is age of acquisition (AoA). The debate on AoA has particularly focused on ultimate attainment and rate of acquisition. More will be said about this in chapter II. Suffice it to say that many scholars agree that - owing to a variety of neurological, affective, cognitive factors as well as to the simple fact of having more time at their disposal in which to acquire the L2 – young learners typically reach higher levels of ultimate attainment. Contrary to what is commonly believed in popular opinion, however, research suggests that young learners do not (always) learn at a faster rate than older learners. Young and older learners are also posited to differ with respect to their aptitude for naturalistic versus instructed learning context – with young learners' cognitive abilities being more suited for a naturalistic context and older learners cognitive characteristics being more suited – or even requiring – formal instruction.

While age of acquisition thus clearly affects rate of acquisition and ultimate attainment and preference for learning setting, it has often been argued that the route of L2 acquisition is unaffected. Research on this topic appears not be as abundant as research on the other effects of AoA, however.

5 Conclusion

In many – otherwise diverse – theories of second language acquisition the acquisition of receptive grammar processing is viewed as preceding the acquisition of productive grammar acquisiton. This view is based on the commonly observed primacy of receptive language skills over productive language skills, and on the logical assumption that a language features needs to be comprehended before it can be acquired for production.

Not surprisingly, given that language production is easier to investigate than language reception skills, SLA acquisition researchers to date have paid almost undivided attention to language production data when seeking to understand the processes involved in L2 acquisition. Driven by the traditional view on language reception, they have furthermore frequently failed to refrain from extrapolating their findings to L2 acquisition in general, thereby implying that the acquisition of receptive L2 grammar knowledge proceeds in the same manner as productive grammar acquisition. This was

the case in the early morpheme acquisition studies as well as in more recent studies investigation the acquisition of grammatical subsystems such as the verb system or syntactic structures. As an additional consequence, neither the theories developed to account for empirical evidence in a research-then-theory approach (e.g. Processability Theory or theories appealing to frequency of occurrence or markedness) nor L2 acquisition theories which offer a (partial) information for the route of L2 grammar acquisition (such as Universal Grammar, Operating Principles or Connectionism) distinguish between language reception and language production. Rather they, too, appear to tacitly assume that what applies to productive grammar development also applies to the acquisition of receptive language skills.

Whether the traditional approach to language reception is justified is far from certain, however. Empirical studies showing the primacy of language production over language reception have led some scholars to claim otherwise. Although they acknowledge that the most common grammar acquisition scenario involves the acquisition of receptive skills before productive skills, some scholars believe that the opposite is also possible. This raises serious doubts concerning the generalization of results from language production studies to language acquisition in general. However, other scholars have argued that the methodology used in the afore-mentioned studies is problematic and consequently do not accept that language production may precede language production.

Even so, there is sufficient reason to take a more cautious stance where the relationship between the acquisition of productive skills and the acquisition of receptive skills is concerned. Until empirical studies shift their focus to the acquisition of receptive language skills, the assumption that the development of receptive grammar knowledge precedes and by extension resembles the development productive grammar knowledge should be considered premature.

Thus, empirical studies which show that the acquisition of productive L2 grammar knowledge is characterized by developmental regularities, both in the order in which grammar features are acquired and in the development of individual L2 grammar feature, need not be taken to suggest that language reception too is marked by systematicity, or if so that the developmental orders and sequences are similar to the ones found in productive L2 grammar acquisition.

II

The Development of Receptive Grammar Knowledge in English as a Second Language: A Cross-Sectional Study

0 Introduction

0.1 Introduction

There is a large body of evidence that the acquisition of productive L2 grammar knowledge is characterised by systematic and possibly universal developmental patterns. A major question that has hitherto remained unsolved however is whether the acquisition of receptive L2 grammar knowledge occurs in a similar fashion as its productive counterpart. Scholarly opinion has recently begun to allow for the possibility that traditional views of the acquisition of receptive language skills as serving, and hence preceding, the acquisition of productive knowledge may be untenable. Therefore there is no reason to assume that the pattern of development of receptive grammar knowledge mirrors that of productive grammar knowledge. Empirical research is needed to gain more insight into the receptive acquisition process and its relation to language production.

Therefore this study presents a cross-sectional examination of the acquisition of receptive grammar knowledge of English as a L2. For the purpose of this study a receptive multiple-choice grammar test which evaluates the knowledge of nine grammatical categories was administered to L2 learners of English.

It is expected that the scope of this study will be too small to allow for strong conclusions. Rather this study should be seen as having an exploratory purpose. In addition to seeking some first insights into receptive grammar acquisition, this study aims just as much at examining the possibilities of future research. For this reason the evaluation of the test materials used and of methodological issues is an integral part of this study.

Finally, the participants in this study are young children (age 6-10) who learn English as an L2 in an immersion setting. Hence this study may also contribute to debates about the role of age in L2 acquisition and about bilingual education. It furthermore forms a

widening of the scope of research into grammar acquisition, since this has hitherto primarily focused on L2 acquisition among adolescent and adult L2 learners of English.

0.2 Early Language and Intercultural Acquisition Studies

The language data that are analysed in this chapter have been collected within the framework of ELIAS (Early Language and Intercultural Acquisition Studies), a Comenius-funded research project which investigates early English immersion education in Europe. ELIAS conducts research in schools across Europe with the aim of gaining insight into the development of the L1, the (receptive) acquisition of L2 grammar and vocabulary knowledge and the development of academic skills and of intercultural awareness. Thus the data analysed in this study will also be used by ELIAS to explore the grammar development of young English immersion pupils across Europe.

In addition, the project has developed research materials such as observation checklists and a grammar test. It is this grammar test which has been used for the purpose of this study. Important to note therefore is that this test was not designed with the purpose of examining receptive grammar development as such. Rather it was developed to gain insight into L2 acquisition in early immersion education.

The advantage of this test is that it is suited for the limited L2 skills of the young L2 learners participating this study. It requires no output and no advanced vocabulary knowledge. Furthermore grammatical categories have been selected which assumed to be acquired in early phases of L2 grammar acquisition. A disadvantage is that the nine grammatical categories of which the receptive knowledge is tested through this test shed only a limited light on the acquisition of receptive grammar knowledge. However, any grammar test which looks at the acquisition process among young L2 learners who are in still in an early acquisition phase will be confronted with this problem.

To conclude, while the grammar test and participants were part of and used with the permission of the ELIAS project, the methodology of this study has been developed independently of ELIAS to suit the specific research purposes of this study.

1 Research Questions and Hypotheses

The aim of this study is to gain insight into the process of acquiring receptive L2 grammar knowledge. In order to achieve this aim, but also taking into account certain specificities of the design of this study, two research questions have been formulated.

A first research question is concerned with the participants' progress (over the course of the first few years of L2 contact) in acquiring receptive knowledge of the L2 grammar.

Notwithstanding the primary focus of this study on systematicity in grammar acquisition, this general research question will be treated first. The reason for this is two-fold. First, it is useful to have some insight into the overall progress made by the participants before analyzing the data for the existence of developmental patterns. Secondly, as explained, this study looks at the L2 grammar acquisition of early L2 learners in, moreover, an immersion setting. Thus this research question may contribute to the scholarly debate concerning age of acquisition and the debates concerning multilingual education. It may be hypothesised that the participants' receptive L2 grammar knowledge will expand as a result of L2 contact.

The second (but in focus primary) research question that will be explored in this study is whether systematicity can be discerned in the order in which the receptive knowledge of the 9 grammatical categories is acquired. If so, it will be further examined how this is related to the developmental patterns found in productive L2 grammar acquisition and the explanations given for these patterns. In view of the complex relationship between receptive and productive language, it is difficult to make any hypotheses with regard to the receptive grammar development in this study.

Put differently, this study is interested in two major aspect of L2 acquisition, namely the rate of receptive grammar acquisition (i.e. research question one) and the route of receptive grammar development (i.e. research question two).

2 Method

2.1 Participants

Seventy-two children who were learning English as an L2 participated in this study. Their age ranged between 75 months and 110 months, with an average of 91.7 months.

Participants were enrolled in the Ecole Cheval Bayard, an early English immersion school in the French Community in Belgium. Approximately 40 percent of the curriculum content (i.e. mathematics, world orientation and creative activities) in the Ecole Cheval Bayard was taught in the L2 English. The remaining 60 percent of the subject matter (reading and writing) was taught in French. The participants' L2 contact thus amounted to approximately 9 hours of English per week. Participants had no contact with the L2 outside of the school (apart from what can be heard through the media).

Age of enrolment in the immersion programme of the Ecole Cheval Bayard lies at 5 to 6 years. None of the pupils selected for this study had any prior knowledge of English upon entering the immersion programme. Consequently, upon entering the immersion programme the participants had passed the turning point, commonly situated at around 4 or 5 years of age (Ortega, 2009b), when the L1 grammar system can be considered acquired and after which all new languages are viewed as consecutively acquired L2s rather than additional, simultaneously acquired L1s.

All participants spoke French as an L1 – some as their only L1 and home language, others as bilinguals in combination with other L1s or home languages. For all participants English was the only L2 (i.e. with age of onset after 4/5 years of age).

Participants were selected on the basis of the time they had spent in the immersion school (and hence their amount of L2 contact). A first group of participants (n=28, average age 82.8 months) was randomly selected from the children enrolled in the first year of primary school. At the time of testing L2 contact for these pupils was 18 months. A second group (n = 23, average age 90.9 months) was randomly selected from pupils enrolled in the second year of primary school. They had spent 29 months in the programme at the time of testing. A third group (n = 21, average age 104.3 months), enrolled in the third year of primary school, had spent 41 months in the immersion programme.

2.2 Instrument

As mentioned, the instrument used in this study is the grammar test (henceforth GT) developed by ELIAS with the aim of investigating the acquisition of L2 English grammar in early English immersion education in Europe. The GT was designed in such

a way that it allows for an assessment of the receptive knowledge of nine grammatical categories. The GT itself is described in section 2.2.1. The nine grammatical categories included in the test are discussed in section 2.2.2.

2.2.1 ELIAS Grammar Test: General Description

In administering the ELIAS GT, the researcher orally presents a test item, in the form of either a word or a sentence containing one of the grammatical categories being investigated, to a participant. The participant is then asked to choose out of three pictures the one which s/he thinks represents the prompt.

Of the three pictures, only the *correct* one corresponds exactly with the test item. For instance, in the case of the prompt *the dog is not sleeping*, the correct picture is a drawing of an awake, sitting dog. A second picture, the *error*, is closely related with the prompt at a semantic level but deviates with regard to the grammatical item that is being investigated. In the example of the above test item which focuses on negation, the '*error* picture' contains a sleeping dog. More specifically, the *error* represents an 'alternative' of the grammatical category that is being investigated. More will be said about these alternatives in section 2.2.2 below. The third picture is a *distractor*, because it is semantically different from (or less related to) the prompt. In the above example, the *distractor* is a picture of a cat. Appendix B contains an example of a *correct*, *error* and *distractor* answer for the test item *cats* (category *plural*).

By incorporating a distractor in the GT, it becomes possible to ascertain whether the participant fails to provide a correct answer because the grammatical item that is being investigated has not been acquired yet (i.e. when an *error* is selected), or because the participant is unable to semantically comprehend the prompt (i.e. when a distractor is selected) – for example due to a restricted vocabulary knowledge. Of course a factor which must not be ignored is that, just as participants may give a correct reply through guesswork, they may be unable to semantically comprehend the prompt but by guessing nonetheless select the '*error* reply'. In case of a '*distractor* reply' however it can be said with certainty that the participant did not process the semantic meaning of the semantic meaning.

The GT test consists of two parts (GT/A and GT/B, administered on different occasions), which are each made up of 27 test items. Thus, every grammatical category

is represented by three test items in each part of the GT. Over the whole test this gives 56 test items (six for every grammatical category).

The GT score sheet can be found in the Appendix.

2.2.2 Grammatical Categories

Nine grammatical categories are included in the ELIAS GT. Table 3 presents grammatical categories along with the abbreviations that will be used in the remainder of this study, and examples.

Table 3 *Grammatical Categories: Abbreviation, Explanation and Example Test Items*

Abbreviation	Grammatical Category	Example
SVO	canonical word order: SVO (subject-verb-object)	The dog is chasing the cat. The cat is chasing the dog.
PLU	plural marker –s: absent/present	Cat Cat**s**
GEN	genitive –s: absent/present	The girl is kissing the boy. The girl is kissing the boy**'s** dog.
PROsubj	personal pronoun, 3rd person singular, subject: masculine/feminine (*he/she*)	**He** is singing. **She** is singing.
PROobj	personal pronoun, 3rd person singular, object: masculine/feminine (*him/her*)	The father is chasing **him**. The father is chasing **her**.
POSS	possessive determiner, 3rd person singular: masculine/feminine (*his/her*)	**His** ball. **Her** ball.
NEG	negation, expressed by the sentence negator *not:* absent/present	The duck is eating. The duck is **not** eating.
VERB	verbal morphology: full verbs, present simple, 3rd person: singular/plural	The fish jump**s**. The fish jump.
COP	verbal morphology: copula verbs, present simple, 3rd person, singular/plural (*is/are)*	The fish **is** black. The fish **are** black.

The majority of grammatical features are morphological. The only syntactic or morphosyntactic categories are SVO and NEG.

For each grammatical category the receptive knowledge of two 'alternatives' of the category is tested. In one part (GT/A of GT/B) of the test, a test item is made up of one 'alternative' of the grammatical category that is being assessed, while in the other part of the test the other alternative is being assessed (by a very similar prompt). The 'alternatives' for each category can be found in the explanation of the test items in Table 3, after the colon. For instance, in the case of GEN the alternatives are *absent/present*, in the case of PROsubj, PROobj and POSS they are *masculine/feminine*. The examples in Table 3 also record the two alternatives of a test item. The other paired test items for each category can be found in the Appendix.

Note further that not only has the test been designed in such a way that a semantic understanding of the test items does not suffice for giving a correct answer (see section 2.2.1 above), the test items have furthermore been made in such a way that the participants must have a receptive knowledge of the grammatical feature being investigated in order to be able to give a correct reply – except in the case of successful guessing.

In the case of VERB and COP, the sentence subjects consist of nouns with a zero plural (sheep, fish and deer). Thus participants cannot rely on their potential knowledge of the noun plural formation to give a correct reply. For the category PLU, the prompts are single words which are either in plural or singular, so that the reverse – namely relying on their potential knowledge of verbal morphology – is not possible.

In the case of the pronouns included in this study (PROobj, PROsubj, POSS) the two alternatives in GT/A and GT/B are based on the distinction feminine/masculine in the 3rd person singular (he/she; him/her; his/her). Hence, in order to be able to select the correct reply, the participants must master the distinction between the masculine and feminine. Conversely, however, an *error* reply may be indicative of either a failure to recognize the distinction masculine/feminine or of an inability to recognize he/she, him/her, his/her as pronouns altogether.

A possible exception is the category GEN, for which there is the possibility that receptive knowledge of the inflected genitive is not strictly required to give a correct reply. For example, in the case of the prompt *The girl is kissing the boy's dog* the presence of the final word *dog* may lead the children to select the picture of a girl kissing a boy's dog rather than the picture of a girl kissing a boy.

It has been claimed that similarity between L1 and L2 items may facility L2 comprehension (Ellis, 2008). Because of this, Berthele et al. (in press) have argued that

> 'it is important (…) to keep an eye on the comparison of structures in the input and output with the other languages present in the individual's multilingual repertoire, in order to find out whether, when and where interlingual transfer could be a reasonable explanation for the patterns discovered in the data'. (p. 4)

Therefore it is worth pointing out that some of the nine grammatical categories in this study have a similar structure in the participants' L1, French, while others have not.

First, similarly to English, canonical word order in French is SVO. By contrast, PLU, POSS and GEN in English are relatively dissimilar from the French variants. Plural *–s* in French, although present in written language, is not pronounced. In selecting the possessive determiner in French, the choice for masculine or feminine is not determined by the gender of the possessor (as in English, e.g. 'her book'), but by the gender of the head noun (e.g. 'son livre'). Furthermore, before being able to make the distinction between masculine and genitive, the participants must recognize 'his' and 'her' as possessive determinatives. The inflected genitive, thirdly, which is very common in English (e.g. John's dog), does not exist in French. Genitive is always expressed by a periphrastic, post-modifying structure (e.g. le chien de Jean) (Biber, Johanson, Leech, Conrad & Finegan, 1999; Grevisse & Goosse, 2007).

For the other categories, the degree of similarity or dissimilarity is less straightforward. In French, personal pronouns in the nominative and accusative case also make a distinction between plural and masculine in the 3rd person singular, but the form of these pronouns is of course different from the English pronouns (*il* and *elle* for *he* and *she*; *le* and *la* for *him* and *her*). Moreover, the position of these pronouns in the sentence is different. Similarly, the copula verb in French has a distinct 3rd person singular (*est*) and plural (*sont*) in the present simple forms but it is unlikely that this will help participants in distinguishing the English forms *is* and *are*. Likewise, the verb formation in the simple present in French for the 3rd person singular (stem + -e) and plural (stem + -ent) is distinct from the English verb formation (stem-s and stem+ø). Finally, the English negator *not* is different in form from the French negator *ne…pas,* and its reduced form. However, once participants have learned that *not* negates the sentence, the use of this negator in English, and its position in the sentence, is similar to the use

and position in French (Biber, Johanson, Leech, Conrad & Finegan, 1999; Grevisse & Goosse, 2007).

2.3 Data Collection Procedure

The GT was administered to each child individually, in a quiet room away from the other children. Administration took approximately 5 to 10 minutes per child, per test part (GT/A or GT/B).

The 2 parts of the GT were administered on different days, with GT/A always being performed first and GT/B being performed on a later day. The time lapse between the administrations of the two tests varied as a result of the school's teaching schedule but was always kept under 8 days.

As part of the ELIAS research project, participant observations had been conducted in the Ecole Cheval Bayard over the course of two years, first in the 3rd year of nursery school and the 1st year of primary school, the second year in the 1st and 2dn year of primary school. As a result, at the time that the GT was conducted the participants enrolled in the 1st and 2nd year of primary school were familiar with the test administrator. This had a positive effect on their anxiety during the test administration.

The oldest groups of participants (3rd year of primary school, 41 months of L2 contact), however, were not part of the ELIAS research project. Furthermore, due to time constraints participant observations for the sake of familiarising the pupils with the researcher were not possible. Nonetheless the majority of the pupils were not conspicuously nervous during the test administration. Even so it is a factor which should not be discarded when analysing the results.

2.4 Design

As explained in section 2.1, participants in this study were selected on the basis of the amount of L2 contact they had received when the GT was administered. Participants can thus be assigned to three groups, henceforth called *L2 contact groups*. Participants in *P1* (enrolled in the first year of primary school) had received 18 months of L2 contact (n=28), *P2* (enrolled in the second year) had received 29 months of L2 contact (n=23) and *P3* (enrolled in the third year) had received 41 months of L2 contact (n=21).

The data gathered in these three L2 contact groups have been used in a cross-sectional investigation of the effect of L2 contact on the level of receptive grammar knowledge of English L2, both in general and more specifically with regard to the 9 grammatical categories included in this study.

Because individual variation is obscured in a cross-section group means and consequently determining developmental patterns is difficult, the participants' GT results will also be examined without classification of the participants into L2 contact groups (namely when conducting implicational scaling, as will be explained below).

2.5 Data Analysis

2.5.1 Effect of Second Language Contact on Receptive Grammar Knowledge

A first research question formulated in this study deals with the overall progress made by (young) L2 learners of English in acquiring receptive knowledge of the L2 grammar. For this purpose, a cross-sectional comparison of the mean scores attained by the L1 contact groups P1, P2 and P3 is conducted, so as to ascertain the effect of L2 contact on the participants' level of receptive grammar knowledge.

The level of receptive grammar knowledge is determined by calculating the percentage of correct responses in GT/A and GT/B. Using *StatPlus:mac LE*, a One-Way between groups Analysis of Variance (ANOVA) followed by post-hoc Tukey HSD analyses are performed in order to determine the statistical significance of the differences in mean scores. Effect size is measured by means of eta squared.

Because the GT tests consisted of two parts that were performed on separate days, the results attained on each of these two parts are analysed in more detail. First, the increases or decreases in the mean scores attained by P1, P2 and P3 on GT/A are analysed, as are the results found for GT/B. In both analyses, One-Way ANOVAs and Tukey-HSD analyses are executed to measure the statistical significance of the results; eta squared is used to express effect size.

Secondly, within each L2 contact group the scores attained on GT/A and GT/B are compared because a high level of consistency in the performance on GT/A and GT/B might point to a more fully required, or more stable, receptive grammar knowledge.

From less consistent results, by contrast, it might be inferred that the participants' receptive grammatical knowledge is less stable, or acquired to a lesser degree. Paired, two-tailed *t*-tests are executed.

Finally, it has been explained that, whereas a correct response signals comprehension of the prompt at the semantic and grammatical level and hence implies receptive acquisition of the L2 grammar, a failure to give the correct response might result either from the participant's lack of receptive knowledge of the grammatical item being investigated (in case of an *error* reply), or from the participant's failure to semantically understand the sentence (in case of a *distractor* reply). A *distractor* reply hence makes it impossible to decide whether the grammatical category is acquired or not.

Therefore, to gauge the impact of semantic comprehension problems on the GT results, the number of *distractor* replies in each L2 contact group will be presented. An ANOVA followed by post-hoc Tukey HSD will show whether the number of *distractors* in one group was larger than in another group.

Secondly, an attempt will be made to predict – or at least roughly estimate – what the GT results for each group would have been if the semantically problematic test items had been semantically understood, and hence participants would have given either a correct or error reply. In order to do this, it will be assumed that the ratio of *correct* and *error* replies within these test items would be the same as the ratio of *correct* and *error* replies already found in the semantically unproblematic test items.

For example, in P1 57.67% of the answers (= mean score) were correct, 34.85% were *errors* and 7.47% were *distractors*. Out of the total of unproblematic replies (92.53%), the correct replies made up 62.33%. It is not unconceivable, then, that if the *distractor* replies were semantically understood, 62.33% of them (= 4.37% out of the total number of replies) would be answered correctly. This would also give a total correct score of 62.33% (correct 57.67% + distractors that would be correct 4.37% = 62.33%). In order words, the L2 contact groups' scores in a GT test free of semantic problems could be estimated simply by ignoring the excluding the *distractor* responses and calculating the % of correct replies out of the total of semantically unproblematic replies in the GT.

It should be emphasised however that the above assumption concerning the ratio of correct and error responses may not be accurate.

2.5.2 Effect of Second Language Contact on the Receptive Knowledge of Nine Grammatical Categories

Following the analysis of the general progress made in acquiring receptive knowledge of the L2 grammar, the remaining part of the study zeroes in on the acquisition of the nine grammatical categories included in the GT, namely SVO, PLU, GEN, PROsubj, PROobj, POSS, NEG, VERB, COP.

For each grammatical category, a cross-sectional comparison is made of the mean scores attained by P1 and P2 and by P2 and P3, to gain insight into the progress, or lack of progress, made in each grammatical category. In addition, the scores for P1 and P3 are compared to determine how much progress is made over a longer period of time. For the statistical analysis One-Way ANOVAs and Tukey HSDs are performed on the mean scores attained by the three L2 contact groups for each grammatical category. In each grammatical category eta squared gives insight into the effect size.

The analysis in this section does not in itself provide an answer to either of the research questions formulated in section. Nevertheless, it is worth doing this analysis for two reasons. On the one hand, it broadens the understanding of the progress made by young immersion pupils learning English as a L2, and as such is an important addition to the overall analysis of the GT scores described in the previous section. Furthermore, it prepares the way for the analyses described in the following sections, the aim of which is to determine the rank order and the scalability (in an implicational scale) of the grammatical categories.

2.5.3 Rank Order of Grammatical Categories Based on the Level of Receptive Grammar Knowledge

The above analytic procedures give insight into the progress made on the GT in general and on the distinct grammatical categories. In a next phase, the mean scores attained for each of the grammatical items are calculated and the categories are ranked for proficiency level.

First, the rank orders found in P1, P2 and P3 will be examined. A One-Way between groups ANOVA and a Tukey HSD are carried out to establish whether the rank order is significant, that is whether the differences in mean scores between the various grammatical categories are significant enough to be able to assign a distinct rank to the

grammatical categories. Next, Spearman rank order correlation coefficient is calculated to ascertain whether the rank orders found for the three L2 contact groups correlate. The coefficient is calculated for P1 versus P2, P2 versus P3 and P1 versus P3.

Because semantic comprehension problems may affect not only the scores attained by the different L2 contact groups on the overall GT test (section 2.5.1) but also the scores attained in each of the grammatical categories, and hence the rank order, the effect of the number of *distractor* replies on the scores attained in the grammatical categories will be calculated. To do this, the same method will be used as when gauging the effect of semantic difficulties on the L2 contact groups' scores (section 2.5.3): ignoring the *distractor* responses and calculating the percentage of correct replies out of the total of semantically unproblematic test items (correct + errors) may – but, importantly, need not – give an acceptable estimation of the scores that would have been attained if all test items had been semantically understood.

A second factor that may influence the rank order lies in the design of the GT, more specifically in the fact that the test consists of paired test items which represent different alternatives of the grammatical categories. As will be explained below, in some categories one alternative is easier than the others, while at the same time not saying much about whether or not the category is acquired. Therefore in each category the scores for the two alternatives will be compared, so as to determine whether the results are representative of the actual level of receptive knowledge of the grammatical categories. For each category, paired, two-tailed *t*-tests (also run in *StatPlus:mac LE*) are used to decide whether the means of the two alternatives are statistically different from each other. For this test too the means (and hence the rank orders) of the participants are not classified into L2 contact groups.

Finally, in many morpheme studies rank orders based on cross-sectional data have been interpreted as acquisition orders. As explained, however, this interpretation has been criticised (chapter I, section 2.1.2). Accordingly, the rank orders found in this study should not be interpreted as evidence of an acquisition order. Rather, a final section contains an implication scaling of the participants' scores.

2.5.4 Developmental Patterns in the Acquisition of Receptive Grammar Knowledge

A method that is frequently used in cross-sectional studies for ascertaining whether systematicity can be found in the development of productive L2 grammar is *implicational scaling* (see also Chapter I, section 2.2.2). Implicational scaling assumes that if systematicity (in the form of developmental patterns) exists, the acquisition of one grammatical item can be used to predict, on the basis of the implicational relationship between the various items, whether or not certain other grammatical items will or will not be acquired.

In binary implicational scaling studies, features are considered 'acquired' when participants have a score of 90% or higher, and 'not acquired' when the score is lower than 90%. Sometimes 80% is taken as the acquisition criterion (Hatch & Farhady, 1982; Rickford, 2002). In an implicational scale, then, grammatical features are ordered from the one that is acquired by the largest number of participants to the one that is acquired by the smallest number of participants. Individual participants' scores (that is, acquired or not acquired) are recorded. If systematicity exists in the acquisition of the items, a grammatical item that is marked as acquired will on one side be flanked by other 'acquired' grammar items, and on the other side by items that have not (yet) been acquired (ibid.).

In this study, an implicational scale is made of the 72 participants' scores on the 9 grammatical items. For each grammatical item, the participants' absolute score on the GT could be 1, 2, 3, 4, 5 or 6 correct replies out of 6 questions. This gives only 6 possible scores in percentage: 16.67, 33.33, 50.00, 66.67, 83.33 or 100. Consequently, using the 90% criterion for acquisition implies that only items with a score of 100% will be marked as acquired. For this reason, and because the 80% is less frequently used than the 90% scale, this study will record both the 90% and the 80% scales.

To determine whether the deviations from the expected implicational pattern of acquired and non-acquired features found in the scale matrix do not invalidate the implicational scale, the *coefficient of reproducibility* (C_{rep}) (also called *Guttman's Index of Reproducibility* (IR), *minimal marginal reproducibility* (MM_{rep}), and most importantly the *coefficient of scalability* are calculated. In some studies only C_{rep} is used to ascertain whether or not the scale matrix is truly suggestive of developmental patterns (Hatch & Farhady, 1982; Rickford, 2002). According to Hatch & Farhady (1982),

however, a scale may have an acceptable C_{rep} but nonetheless turn out *not* to be scalable when the *coefficient of scalability* is calculated.

Finally, it must be noted that implicational scales have thus far never been made with the type of grammar test used in this study. This study therefore should only be seen as a tentative, exploratory use of implicational scaling.

3 Results

3.1 Effect of Second Language Contact on Receptive Grammar Knowledge

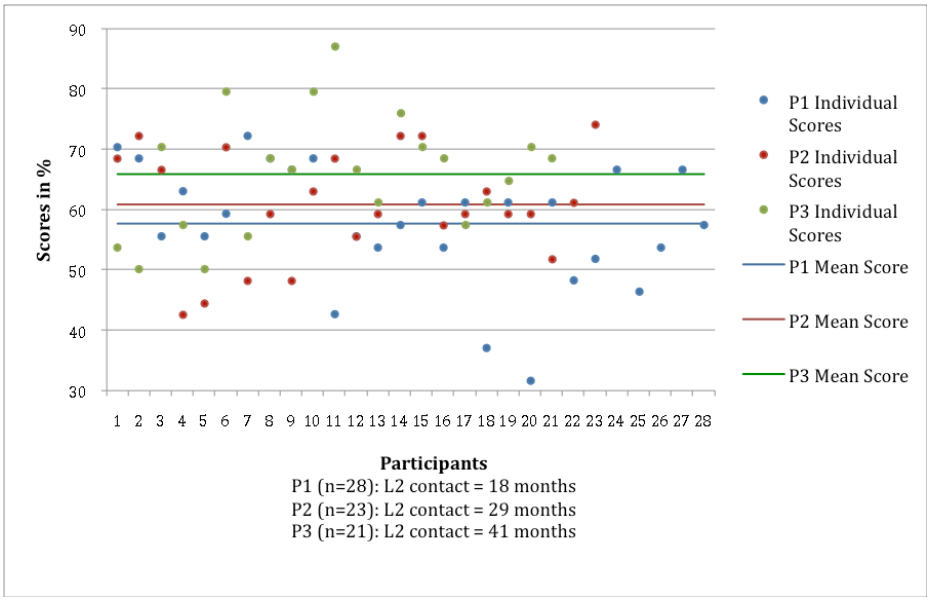
The first research question of this study is concerned with the general progress in the acquisition of receptive L2 grammar knowledge by young learners of English as a L2. For this purpose, the overall GT scores attained by participants in P1 (n = 28; L2 contact = 18 months), P2 (n = 23; L2 contact = 29 months) and P3 (n = 21; L2 contact = 41 months) are compared.

First, the results for the GT in total are analysed (section 3.1.1). Next, the results for GT/A and GT/B are compared (section 3.1.2). Finally, the role played by semantic comprehension. is examined (section 3.1.3)

3.1.1 Receptive Knowledge of Second Language Grammar: General Results

Figure 1 presents the individual and mean GT scores (in %) attained by P1, P2 and P3.

Figure 1 *Individual and Mean GT Scores (P1, P2 and P3)*



First, Figure 1 reveals that the mean score attained in P1 was nearly 60%. Given that the participants in P1 had had only had 18 months of L2 contact, in an immersion setting, at the time of testing, this can be considered a relatively high score.

Secondly, the graph displays a very modest rise in the level of receptive grammar knowledge attained by P2 in comparison to P1 (from somewhat less than 60% to just over 60%) and a somewhat larger but still modest increase between P2 and P3 (with P3 attaining a mean score of around 65%).

Thirdly, a high degree of individual variation is evident in Figure 1, with some participants in P1 attaining a higher level of receptive grammar knowledge than certain participants in P2 and P3, in spite of an additional 11 and 23 months of L2 contact at the time of testing for P2 and P3 respectively. Likewise, some children in P2 attained a higher score than certain children in P3, who had had additional 12 months of L2 contact.

To determine the significance of the increases observed in the L2 contact groups' mean scores, a One-Way ANOVA was performed. The test revealed a significant effect of L2 contact on the level of receptive grammar skills at the $p < .05$ level of significance; $F(2, 69) = 4.28$, $p = .02$. Further, eta squared indicated a fairly high effect size; $\eta^2 = .11$.

A post-ANOVA Tukey HSD test further disclosed that the mean score for P1 ($M = 57.67$, $SD = 10.00$) was significantly different from P3 ($M = 65.87$, $SD = 9.89$); $p < .05$, so that it can be concluded that the participants' receptive knowledge of the L2 English grammar increases over the course of two years. Importantly, however, neither the difference in the mean scores of P1 and P2 ($M = 60.71$, $SD = 9.22$) nor the difference in the mean scores of P2 and P3 was found to be significant. Thus, no conclusions can be drawn with respect to the participants' grammar acquisition over the course of one year, nor can anything be surmised with regard to the larger increase in mean scores found in the P2-P3 comparison than in the P1-P2 comparison.

3.1.2 GT/A and GT/B Compared

As mentioned above, the GT consisted of two parts (GT/A and GT/B) which were administered on different days. For this reason the scores for GT/A and GT/B will be examined in more detail.

This will be done in two ways. First, the scores attained by the L2 contact groups will be analysed for GT/A and GT/B seperately. Secondly, the difference between GT/A and GT/B within each L2 contact group group will be compared.

The following picture presents the means scores (in %) for P1, P2 and P3 on GT/A (left, darker blue columns) and GT/B (right, lighter blue columns).

Figure 2 *Mean Scores GT/A and GT/B in P1, P2, P3*

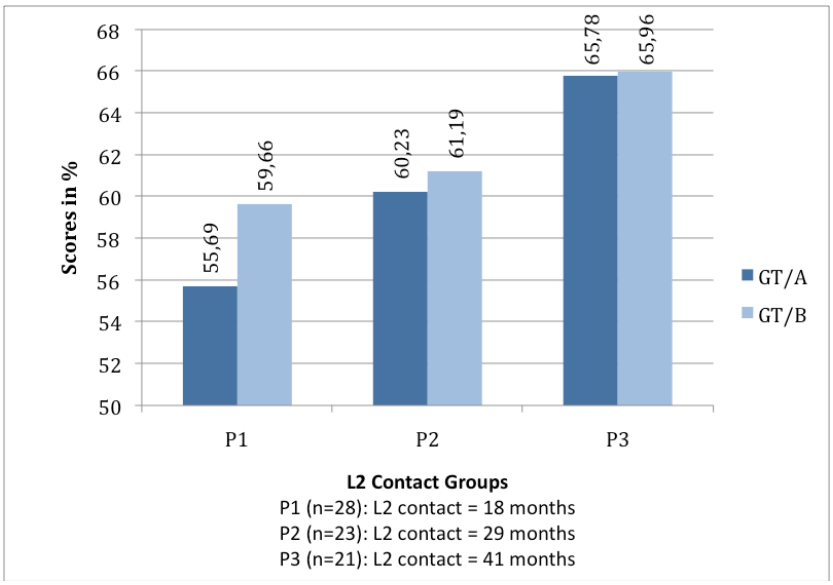


Figure 2 shows an increase in scores as a result of L2 contact (P1, P2 and P3) both for GT/A and GT/B. However the increases are not equal in size.

An ANOVA run on the GT/A result confirmed that L2 contact had a significant effect on the participants' level of receptive grammar proficiency; $F(2, 69) = 4.39$, $p = .02$. Furthermore, η^2 was found to be .11, indicating a fairly large effect size. A Tukey HSD test showed that there was a significant difference between P1 ($M = 55.69$, $SD = 12.37$) and P3 ($M = 65.78$, $SD = 11.11$); p. < .05. However, neither the difference between P1 and P2 ($M = 60.23$, $SD = 11.72$) nor the difference between P2 and P3 was significant. This finding corresponds with the findings for the overall GT scores reported in section 3.1.1. For GT/B, interestingly, the ANOVA revealed no significant effect of L2 contact on the receptive grammar knowledge; $F(2, 69) = 2.60$, $p = 0.08$. Effect size was low, with $\eta^2 = .07$. The reason for this will become clear below.

Concerning the difference between the scores attained on GT/A and GT/B by the three L2 contacts groups, it may be observed that for all three L2 contacts groups the score for GT/B was higher than the score for GT/A. In the case of P3, however, the difference in scores between GT/A and GT/B is negligible (amounting to only + 0.18). For P2, a somewhat larger but still inconsiderable difference (+ 0.96) between GT/A and GT/B is discernible, while P1 shows the most substantial difference in mean scores (+ 3,97).

In the case of P1, a paired *t*-test showed a statistical significance between GT/A (M = 55.69, SD = 12.37) and GT/B (M = 59.66, SD = 9.58); $t(27) = 2.21$, p = 0.03. Not surprisingly, in P2 the minor difference in mean scores between GT/A (M = 60.23, SD = 11.23) and GT/B (M = 61.19, SD = 9.40) was not statistically significant; $t(22) = .44$, $p = .66$. Neither did a *t*-test reveal the negligible difference between GT/A (M = 65.78, SD = 11.11) and GT/B (M = 65.96, SD = 10.45) in P3 to be statistically significant; $t(20) = .09$, $p = .93$.

In short, in P2 and P3 the scores on the GT may not be strikingly better than in P1 but performance on the two parts do appear to be more consistent, with a similar score on GT/A and GT/B. This might suggest that the degree to which the grammatical items in this test are acquired is higher in the case of P2 and P3 than in the case of P1. Because the grammar knowledge in P1 is not yet as stable as in P2 and P3, on one occasion (GT/B) the participant is able to correctly process the grammatical item while on another occasion the grammatical item is not processed (correctly) (GT/A). By extension, on one occasion (GT/B) the participants in P1 are able to attain a score which is not significantly lower than the score attained by the participants in P2 and P3. On a different occasion (GT/A), their score does appear to be significantly lower. This explanation cannot be more than speculation, however.

An alternative explanation is that the scores on GT/B are higher because the pupils have been familiarised with the test and as a consequence of this are less anxious during the test administration. The younger age of the pupils in P1 may cause their performance to be more susceptible to these factors. Or this explanation may work alongside the previous one, in the sense that the less stable grammar knowledge of the participants in P1 is more easily affected, in a negative way, by factors such as anxiety.

3.1.3 The Role of Semantic Comprehension Difficulties

As explained, in performing the GT participants could choose between a *correct* picture (correspondence at the semantic and grammatical level), a similar but incorrect picture (semantically related, but no grammatical correspondence) and a *distractor* (no semantic relation). An incorrect response, or *error,* suggests that the participant understood the overall semantic content of the sentence but did not possess the receptive grammatical knowledge required to recognize the correct picture. Opting for the *distractor* signifies that the participant did not understand the sentence even at a superficial propositional level so that it cannot not be ascertained whether the grammatical category was acquired or not. Thus, to be able to compare the GT scores the percentage of *distractor* answers would have to be constant over the three groups.

Table 4 therefore shows the mean number of *distractor* answers and *error* answers, expressed in % out of the total (i.e. incorrect + correct) selected by the participants in P1, P2 and P3.

Table 4 *The Role of Distractors in P1, P2, and P3*

	P1	P2	P3
Correct (% out of total)	57.67	60.71	65.87
Errors (% out of total)	34.85	34.14	28.40
Distractors (% out of total)	7.47	5.15	5.73
Total Incorrect (% out of total)	*42.33*	*39.29*	*34.13*

As can be seen in Table 4, the number of *distractor* replies was highest in P1 (M = 7.47, SD = 6.14) and lowest in P2 (M = 5.15, SD = 7.47). Surprisingly the participants in P3 appeared to experience more difficulties in semantically processing the test items (M = 5.73, SD = 8.05) than in P2. However an ANOVA showed no statistical significance; $F(2,69) = .73$, $p = .48$.

An alternative way of judging the influence of semantic processing is by looking at the percentage of correct replies out of the total of semantically unproblematic test items, rather than out of the total of semantically unproblematic *and* problematic replies. As explained in section 2.5.3, it is conceivable that the ratio of *correct* responses to *error* responses within the *problematic* test items would be the same as the ratio of *correct*

responses to *error* responses found within the *unproblematic* test items. Hence calculating the L2 contact groups' scores by discarding the distractor responses might be a valid prediction of the scores in a semantically unproblematic GT.

Mean scores for the three L2 contact groups when discarding *distractor* responses are: 62.0% for P1 (SD = 70.11), 63.9% for P2 (SD = 52.05) and 69.9% for P3 (SD = 68.42). An ANOVA conducted on these scores showed high statistical significance; $F(2,69) = 6.05$, $p = .004$. Post-hoc Tukey HSD further showed that the decrease in scores between P1 and P2 was insignificant. Both the mean score difference between P2 and P3 ($p < .01$) and between P1 and P3 ($p < .05$) were shown to be significant, however - the former as a result of the larger number of *distractor* responses in P3 in comparison to P2.

As there is no certainty that the ratio of correct and error responses within the problematic test items would be as predicted above, the scores found when discarding *distractor* replies cannot replace the scores that are found when counting all *distractor* replies as incorrect. It may be clear from the above analysis however that the role of semantic processing problems should not be disregarded.

3.2.4 Summary

In summary, the previous section has shown a statistically significant increase in scores between P1 and P3, from a mean score of 57.67 % to a mean score of 65.87% (+8.2). By contrast neither the increase in mean score between P1 and P2 (+ 3.04, from 57.67 % to 60.71 %) nor the increase between P2 and P3 (+ 5.17, from 60.71% to 65.87%) was statistically significant. Given that the participants in P1 had had only 18 months of L2 contact, the score of P1 can be deemed high. The (rather small) increase in scores between P1 and P3 indicates that the immersion pupils' receptive grammar knowledge further expands over the course of 2 years, although at a much slower rate. Semantic comprehension problems may give a somewhat distorted picture of the GT results however.

Finally, it was found that in P1, the scores on GT/A and GT/B – the two paired subparts of the test – were significantly different, with the scores on GT/B being higher than the scores on GT/A. In P2 and P3, the minor differences between P1 and P2 were not significant. This may suggest – though it must be conceded that this can only be mere

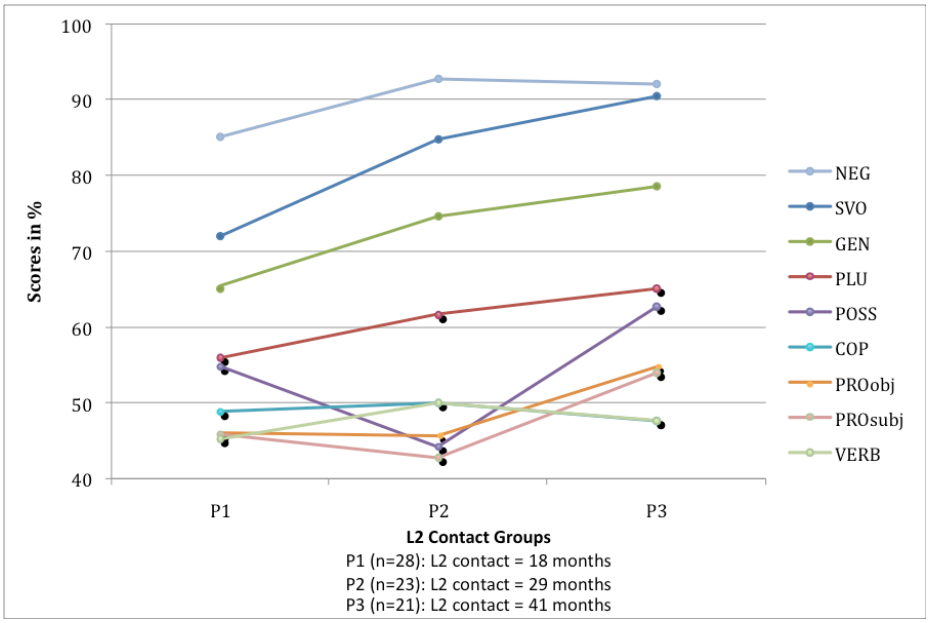
speculation at this stage – that the receptive grammar knowledge in P2 and P3 is not only more expanded but is also more stable than the receptive grammar knowledge in P2.

3.2 Effect of Second Language Contact on the Receptive Knowledge of Nine Grammatical Categories

3.2.1 General Overview of Findings

In this section the results attained in the various grammar categories are examined in more detail. Figure 3 contains a graphic presentation of the mean scores (in %) attained by the L2 contact groups P1, P2 and P3 in each of the 9 grammatical categories.

Figure 3 *Mean Scores for P1, P2 and P3 per Grammatical Category*



The general picture presented by Figure 3 is as follows. For certain grammatical categories (namely SVO, GEN, PLU) a (variegated) rise in scores can be observed both when comparing P1 and P2 and when comparing P2 and P3. This then results in an increased grammar knowledge when comparing P1 and P3. However the rate at which these categories are acquired appears to vary considerably from one category to another.

For other categories a more irregular progression or a lack of progression is evident, characterised by either an increase in score from P1 to P2 followed by a decrease from P2 to P3 (NEG, COP, VERB) or conversely a decrease from P1 to P2 followed by an

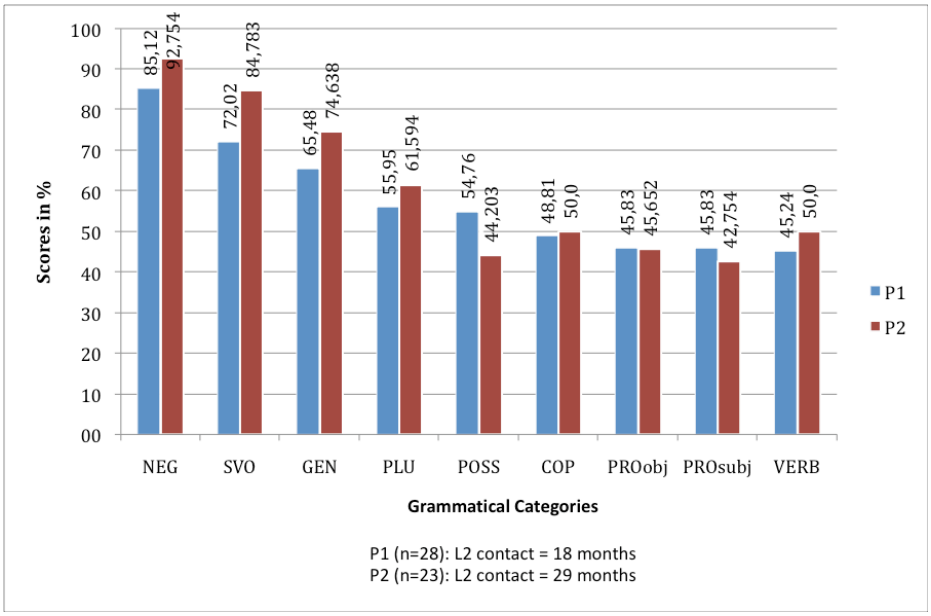
increase from P2 to P3 (POSS, PROobj, PROsubj). For one category the gross result is a decrease in score (COP), for the other categories the overall result is an improved knowledge.

However, in view of the poor statistical significance of the overall grammar progress (see previous sections), many of the rises and falls in scores for the individual categories may not be significant. Therefore the results of Figure 3 will be examined in more detail below. A statistical analysis will be made of the increases and decreases in score from P1 to P2 (section 3.2.2), from P2 to P3 (section 3.2.3), and from P1 to P3 (section 3.2.4).

3.2.2 Cross-Sectional Comparison of Receptive Grammar Knowledge in P1 and P2

Figure 4 displays the scores (%) attained by participants in P1 (blue) and P2 (red) in the nine grammatical categories. Categories are ranked for proficiency level in P1.

Figure 4 *Grammatical Categories: Comparison of Scores in P1 and P2.*



When comparing the scores attained by P1 and P2 in this cross-sectional study, an increase in the mean scores can be observed for the following 8 categories: NEG (+7.63), SVO (+12.76), GEN (+9.16), PLU (+ 5.64), COP (+1.19), and VERB (+4.76). A drop in scores, conversely, can be attested for the category POSS (-10.56) and to a much smaller extent for PROsubj (-3.08) and PROobj (-0.18)

ANOVAs were conducted on the data gathered in P1, P2 and P3 for each of the grammatical categories. Where the ANOVA indicated a significant difference between P1, P2 and P3, a post-hoc Tukey HSD was performed to determine the significance of the difference in mean score between P1 and P2, between P2 and P3 and between P1 and P3.

An effect of L2 contact on the level of receptive grammar knowledge was *not* found for the following 6 categories: NEG ($F(2, 69) = 1.91$, $p = .15$), GEN ($F(2, 69) = 2.75$, $p = .07$), PLU ($F(2, 69) = 1.21$, $p = .30$), COP ($F(2, 69) = 0.10$, $p = .90$), PROobj ($F(2, 69) = 1.27$, $p = .29$), PROsubj ($F(2, 69) = 1.59$, $p = 0.21$) and VERB ($F(2, 69) = 0.48$, $p = 0.62$). Further, eta squared showed small to modest effect sizes for all 6 categories (NEG: $\eta^2 = .05$; GEN: $\eta^2 = .08$; PLU: $\eta^2 = .03$; COP: $\eta^2 = .003$; PROobj: $\eta^2 = 0.4$; PROsubj: $\eta^2 = .04$ and VERB: $\eta^2 = .01$).

That being the case, there were only 2 categories with a significant effect of L2 contact on receptive grammar knowledge. The ANOVA for SVO suggested a statistical significance at the $p < .05$ level; $F(2, 69) = 3.57$, $p = .03$. For POSS, the statistical significance bordered at the $p < .01$ level of significance; $F(2, 69) = 4.87$, $p = .01$. Effect size for SVO was determined as medium; $\eta^2 = .09$. For the category POSS, eta squared showed a larger effect size; $\eta^2 = .12$.

However a Tukey HSD disclosed that neither the increases in mean scores observed in the category SVO between P1 ($M = 72.02$, $SD = 30.78$) and P2 ($M = 84.78$, $SD = 21.85$) nor the decline in mean score in the category POSS between P1 ($M = 54.76$, $SD = 19.17$) and P2 ($M = 44.20$, $SD = 17.84$) was significant.

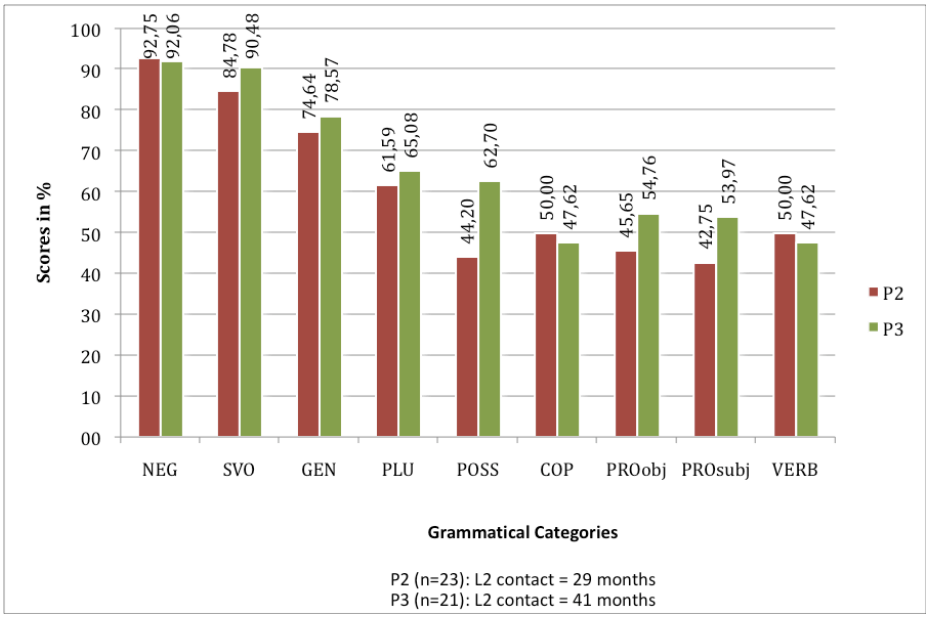
In short, although Figure 3 and Figure 4 suggested an increased receptive grammar knowledge for the categories NEG, SVO, GEN, PLU, COP and VERB and conversely a reduced grammar knowledge for POSS, PROsubj and PROobj, none of the differences in mean scores between P1 and P2 was found to be statistically significant. Effect sizes were small to medium, with only POSS having a somewhat larger effect size of $\eta^2 = .12$.

3.2.3 Cross-Sectional Comparison of Receptive Grammar Knowledge in P2 and P3

In this chapter, the above analysis will be repeated for the differences in mean scores observed when comparing P2 and P3.

Figure 5 presents the mean scores (in %) for the nine grammatical categories, as attained by the participants in P2 (red) and in P3 (green).

Figure 5 *Grammatical Categories: Comparison of Scores in P2 and P3*



Disregarding statistical results for a moment, the increase in mean scores from P1 to P2 was continued into P3 only for the grammatical categories SVO (+5.69), GEN (+3.93) and PLU (+3.49). Contrastingly, a drop in score is discernable for the categories NEG (-0.69), COP (-2.38) and VERB (-2.38).

In addition, POSS (+10.56), PROobj (9.11) and PROsubj (11.21), the three categories which were marked by a fall in score from P1 to P2, now show a rise in score.

As explained in section 3.2.2, a One-Way ANOVA showed a significant effect of L2 contact only in the categories SVO and POSS. The score differences that can be discerned in the other categories are not significant. Moreover, a Tukey HSD performed on the data for SVO indicated no significant difference in mean score between P2 (M = 89.13, SD = 22.25) and P3 (M = 90.48, SD = 18.69).

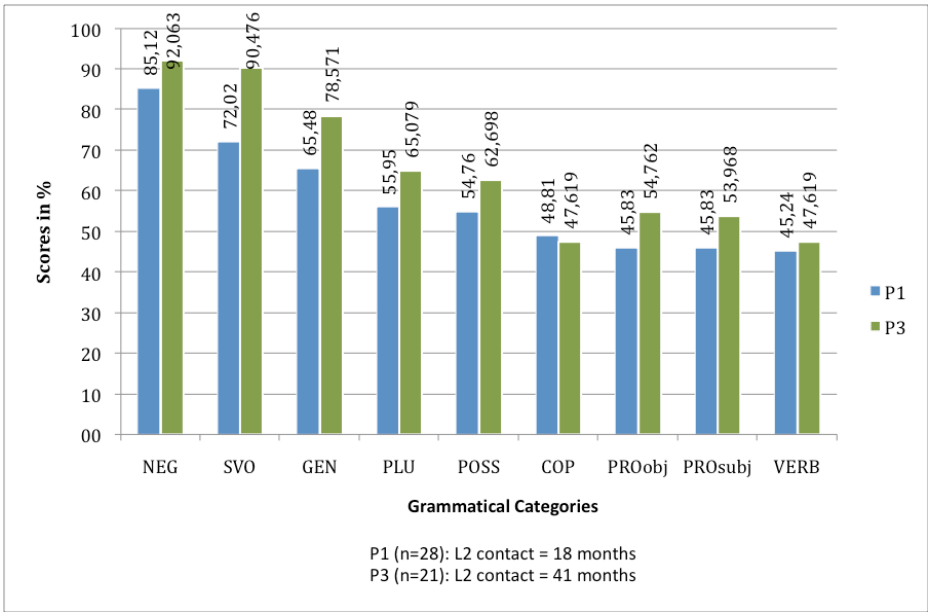
However for the category POSS the rise in score between P2 (M = 44.20, SD = 17.84) and P3 (M = 62.70, SD = 22.30) was found to be significant by Tukey HSD; $p < .05$.

In short, except for the category POSS no conclusions can be reached with regard to the progress of lack of progress evident in the other categories.

3.2.4 Cross-Sectional Comparison of Receptive Grammar Knowledge in P1 and P3

To recapitulate, Figure 6 displays the scores (expressed in %) attained by P1 and P3, thus showing the progress made over the course of 2 years of L2 contact in an immersion setting. Given that the discussion of the overall GT results (section 3.1) showed a statistical significant effect of L2 contact only over the course of 2 years (P1-P3), the results for the distinct grammatical categories might also be more significant when comparing P1 and P3 than when comparing P1 and P2 or P2 and P3, as has been done above (sections 3.2.2 and 3.2.3).

Figure 6 *Grammatical Categories: Comparison of Scores in P1 and P3*



It has already been observed above (section 3.2.4, Figure 3) that all categories except COP are marked by an increase in score in P3 in comparison to P1. For COP a small decrease was found. Furthermore, it has been explained in sections 3.2.2 and 3.2.3 that ANOVAs performed on the score attained by P1, P2 and P3 in the nine grammatical categories revealed a significant effect of L2 contact on the level of receptive grammar knowledge only for the categories SVO and POSS.

With regard to the effect of L2 contact over the course of two years, then, a post-hoc Tukey HSD revealed a statistically significant difference between P1 (M = 72.02, SD = 30.78) and P3 (M = 90.48, SD = 18.69) for the category SVO ($p < .05$) only.

3.2.5 Summary

In short, while all categories except COP show a rise in score between P1 and P3, the picture is more complicated for the intervening phases P1 – P2 and P2 – 3. In comparing P1 and P2, a rise in score can be attested for the categories NEG, SVO, GEN, PLU, COP, VERB. By contrast, POSS, PROobj and PROsubj show a fall in mean score. Of the former, SVO, GEN and PLU continue to show an increase when comparing P2 and P3, while the scores for NEG, COP and VERB decrease in this second phase. POSS, PROobj and PROsubj compensate for their initial drop in score by a larger increase between P2 and P3.

With the exception of the rise in score attested in the category POSS when comparing P2 and P3 and the rise in score attested in the category SVO when comparing P1 and P2, no statistical significance was found. This renders it impossible to draw any valid conclusions with regard to the rate of acquisition for the nine categories and the relative difficulty experienced by L2 learners in acquiring these categories.

3.3 Rank Order of Grammatical Categories Based on Levels of Receptive Knowledge

3.3.1 Rank Orders Based on Mean Scores: A Comparison of P1, P2 and P3

In P1, ranking the categories for the participants' level of receptive knowledge gives the following result: 1 NEG, 2 SVO, 3 GEN, 4 PLU, 5 POSS, 6 COP, 7/8 PROobj and PROsubj, 9 VERB. Participants had attained the highest mean score on the category that is ranked first; the lowest score on the category that is ranked last.

An ANOVA performed on the scores of these nine categories revealed a high degree of statistical significance; $F(8, 243) = 12.04$, $p < .001$. Eta squared further showed a large effect size of .28.

According to a post-hoc Tukey HSD, however, the difference in mean score (in %) between two categories had to be larger than 17.78 in order to be significant at the $p < .05$ level of significance and larger than 20.54 in order to be significant at the $p < .01$ level of significance.

Table 5 shows which mean scores are significantly different from one another other according to the Tukey HSD result. Mean scores and Standard Deviations are also recorded.

Table 5 *Rank Order P1: Significance of Mean Score Differences*

	NEG *M* = 85.12 *SD* = 19.43	SVO *M* = 72.02 *SD* = 30.87	GEN *M* = 65.48 *SD* = 21.72	PLU *M* = 55.95 *SD* = 21.38	POSS *M* = 54.76 *SD* = 19.17	COP *M* = 48.81 *SD* = 15.67	PROobj *M* = 45.83 *SD* = 22.05	PROsubj *M* = 45.83 *SD* = 16.74	VERB *M* = 45.24 *SD* = 20.21
NEG *M* = 85.12 *SD* = 19.43									
SVO *M* = 72.02 *SD* = 30.87	*n/s*								
GEN *M* = 65.48 *SD* = 21.72	**p < .05**	*n/s*							
PLU *M* = 55.95 *SD* = 21.38	**p < .01**	*n/s*	*n/s*						
POSS *M* = 54.76 *SD* = 19.17	**p < .01**	*n/s*	*n/s*	*n/s*					
COP *M* = 48.81 *SD* = 15.67	**p < .01**	**p < .01**	*n/s*	*n/s*	n/s				
PROobj *M* = 45.83 *SD* = 22.05	**p < .01**	**p < .01**	**p < .05**	*n/s*	*n/s*	*n/s*			
PROsubj *M* = 45.83 *SD* = 16.74	**p < .01**	**p < .01**	**p < .05**	*n/s*	*n/s*	*n/s*	*n/s*		
VERB *M* = 45.24 *SD* = 20.21	**p < .01**	**p < .01**	**p < .05**	*n/s*	*n/s*	*n/s*	*n/s*	*n/s*	

On the basis of Table 5, the following conclusions can be drawn.

First, at 18 months of L2 contact the participants' receptive knowledge of the category NEG (first column) was better than their knowledge of all other categories, except for SVO. Secondly, the participants' receptive knowledge of SVO (second column) was significantly better than the receptive knowledge of COP, PROobj, PROsubj and VERB. Thirdly, the knowledge of GEN (third column) was significantly better than that of PROobj, PROsubj and VERB.

The remaining mean score differences were found not to be significant.

When comparing P1 and P2, next, it was observed that the rank order found for P1 did not correspond exactly with the rank order observed for P2. The order of the 4 highest categories (1 NEG, 2 SVO, 3 GEN, 4 PLU) remained unchanged. In the 5 lower-ranked categories, the order changed to 5/6 COP and VERB, 7 PROobj, 8 POSS, 9 PROsubj.

An ANOVA conducted on the data from P2 suggested that the differences between the scores attained on the grammatical categories were highly significant; $F(8, 198) = 22.89$, $p < .001$. Eta squared showed a high effect size; $\eta^2 = .48$.

Post-hoc Tukey HSD revealed that the difference in mean score (in %) between two categories had to exceed 17.46 in order to be significant at the $p < .05$ level of significance and 20.16 in order to be significant at the $p < .01$ level of significance.

Table 6 presents the statistical significance of the differences in mean score between the nine categories. Note that because of the different rank orders, the Tables for P1 and P2 cannot be compared simply by placing them next to each other.

Table 6 *Rank Order P2: Significance of Mean Score Differences*

	NEG *M* = 92.75 *SD* = 12.13	SVO *M* = 84.78 *SD* = 21.85	GEN *M* = 74.64 *SD* = 21.24	PLU *M* = 61.59 *SD* = 19.74	COP *M* = 50.00 *SD* = 17.41	VERB *M* = 50.00 *SD* = 15.08	PROobj *M* = 45.65 *SD* = 18.95	POSS *M* = 44.20 *SD* = 15.08	PROsubj *M* = 42.75 *SD* = 23.48
NEG *M* = 92.75 *SD* = 12.13									
SVO *M* = 84.78 *SD* = 21.85	*n/s*								
GEN *M* = 74.64 *SD* = 21.24	**p < .05**	*n/s*							
PLU *M* = 61.59 *SD* = 19.74	**p < .01**	**p < .01**	*n/s*						
COP *M* = 50.00 *SD* = 17.41	**p < .01**	**p < .01**	**p < .01**	*n/s*					
VERB *M* = 50.00 *SD* = 15.08	**p < .01**	**p < .01**	**p < .01**	*n/s*	n/s				
PROobj *M* = 45.65 *SD* = 18.95	**p < .01**	**p < .01**	**p < .01**	*n/s*	*n/s*	*n/s*			
POSS *M* = 44.20 *SD* = 15.08	**p < .01**	**p < .01**	**p < .01**	*n/s*	*n/s*	*n/s*	*n/s*		
PROsubj *M* = 42.75 *SD* = 23.48	**p < .01**	**p < .01**	**p < .01**	**p < .05**	*n/s*	*n/s*	*n/s*	*n/s*	

In comparison to the rank order in P1, slightly more statistically significant differences can be observed.

In the 6 lowest-ranked categories, the difference between PLU (the highest-ranked category of the 6) and PROsubj (now the lowest-ranked category) in P2 was significant at the level of $p < .05$. Similarly to the situation in P1, no other statistically significant differences can be found.

For the other categories, minor changes can be discerned. The mean score of SVO was no longer only significantly different from PROobj, PROsubj VERB and COP. Owing to the fall in the mean score attained in the category POSS (-10.56), along with the increase in the score for SVO (+12.76), the difference between SVO and POSS in P2 was also significant in P2. Further, due to the relatively high increase in the mean score

of SVO (+12.76), together with a much smaller increase in the score of PLU (+5.64), the mean score difference between these two categories also became significantly different in P2.

Finally, in P2 the level of significance of the difference in mean score between on the one hand NEG and on the other hand GEN, PLU, COP, VERB, PROobj, POSS and PROsub, as well as the insignificance of the difference between NEG and SVO and GEN and PLU remained unchanged in comparison to P1.

The rank order that was found in P3 constituted a partial return to the rank order found in P1, in that POSS resumes the 5th rank. On the other hand, some changes can be discerned in the 6 lowest-ranked categories. The lowest mean scores were attained in the categories VERB and COP. The mean scores attained in PROobj and PROsubj were somewhat higher.

An ANOVA conducted on the differences in mean score found in P3 showed highly statistical significance; $F(8, 180) = 15.94$, $p < .001$. Effect size, further, was revealed to be high; $\eta^2 = .41$. A post-hoc Tukey HSD, however, showed that in order to be significant at the $p < .05$ level of significance, inter-categorical differences had to be larger than 19.33. To reach a $p < .01$ level of significance the difference had to exceed 22.32. Table 7 gives an overview of the statistical significance of all inter-categorical differences.

Table 7 *Rank Order P3: Significance of Mean Score Differences*

	NEG *M* = 92.06 *SD* = 12.49	SVO *M* = 90.48 *SD* = 18.69	GEN *M* = 78.57 *SD* = 16.79	PLU *M* = 55.95 *SD* = 21.38	POSS *M* = 62.70 *SD* = 22.30	PROobj *M* = 54.76 *SD* = 24.23	PROsubj *M* = 53.97 *SD* = 24.67	VERB *M* = 47.62 *SD* = 15.17	COP *M* = 47.62 *SD* = 19.92
NEG *M* = 92.06 *SD* = 12.49									
SVO *M* = 90.48 *SD* = 18.69	*n/s*								
GEN *M* = 78.57 *SD* = 16.79	*n/s*	*n/s*							
PLU *M* = 65.08 *SD* = 21.02	**p < .01**	**p < .01**	*n/s*						
POSS *M* = 62.70 *SD* = 22.30	**p < .01**	**p < .01**	*n/s*	*n/s*					
PROobj *M* = 54.76 *SD* = 24.23	**p < .01**	**p < .01**	**p < .01**	*n/s*	n/*s*				
PROsubj *M* = 53.97 *SD* = 24.67	**p < .01**	**p < .01**	**p < .01**	*n/s*	*n/s*	*n/s*			
VERB *M* = 47.62 *SD* = 15.17	**p < .01**	**p < .01**	**p < .01**	*n/s*	*n/s*	*n/s*	*n/s*		

COP *M* = 47.62 *SD* = 19.92	**p < .01**	**p < .01**	**p < .01**	*n/s*	*n/s*	*n/s*	*n/s*	*n/s*

As was the case in P1 and also in P2 (except for PLU - PROsubj in the latter L2 contact group), the difference between the six lowest-ranked categories was not statistically significant. Consequently the rank order that was found for these categories was meaningless.

For the other three categories (NEG, SVO, GEN) the rank order found in P3 was marked by minor changes in comparison to P2 and/or P1. The differences in mean score between SVO and PLU and between SVO and POSS, which were insignificant in P1 but significant in P2, remained significant ($p < .01$). The relationship between GEN and PLU remained marked by an insignificant difference, while the mean score difference between GEN and COP (marked by insignificance in P1 and significance in P2) was significant in P3. The difference in mean score between GEN and POSS, which was insignificant in P1 but significant in P2, became insignificant again. Finally, in P3 the difference in mean score between NEG and GEN returned to the state of P1, i.e. statistical insignificance.

To summarise the complex and detailed discussion presented above, Table 8 below shows which inter-categorical mean score differences remained significant across P1, P2 and P3 (shaded in blue), and for which differences, by contrast, statistical significance varied (shaded in grey).

The rank order in this Table is the one found in P1 – a random choice. Differences in the *level* of statistical significance (i.e. $p < .05$ or $p < .01$) are not marked as differences. Where the statistical significance varied, both *n/s* and the level of significance ($p < .01$ or $p < .05$) are recorded.

Table 8 *Significance of Mean Score Differences: Overview P1, P2 and P3.*

Shaded in grey: statistical significance varies across P1, P2 and P3
Shaded in blue: statistical (in)significance is constant across P1, P2 and P3.

	NEG	SVO	GEN	PLU	POSS	COP	PROobj	PROsubj	VERB
NEG									
SVO	*n/s*								
GEN	n/s p < .05	*n/s*							
PLU	***p* < .01**	*n/s* *p* < .01	*n/s*						
POSS	***p* < .01**	*n/s* *p* < .01	*n/s* *p* < .01	*n/s*					
COP	***p* < .01**	***p* < .01**	*n/s* *p* < .01	*n/s*	n/s				
PROobj	***p* < .01**	***p* < .01**	***p* <** **.05/.01**	*n/s*	*n/s*	*n/s*			
PROsubj	***p* < .01**	***p* < .01**	***p* <** **.05/.01**	*n/s* *p* < .05	*n/s*	*n/s*	*n/s*		
VERB	***p* < .01**	***p* < .01**	***p* <** **.05/.01**	*n/s*	*n/s*	*n/s*	*n/s*	*n/s*	

3.3.2 Correlation between the Rank Orders

To determine the correlation between the rank orders found in P1, P2 and P3, Spearman's rank order correlation coefficient was used. Given that not many differences in rank orders were found above, it is not surprising that a high degree of correlation was found in all three cases; P1 - P2: ϱ = .80; P2 - P3 : ϱ = .73; P1 - P3: ϱ = .93.

However, it must be taken into account that, in order for the Spearman rank order coefficient to be truly meaningful, the differences between the rank and consequently the orders themselves must be meaningful as well (Hatch & Farhady, 1982). However, as the ANOVA and post-hoc Tukey HSDs in the previous section (3.3.1) have shown, this was not the case for all grammatical items. Consequently, the high rank order correlation coefficients found in this study may not be meaningful.

3.3.3 Overall Rank Order and the Role of Semantic Comprehension

This section examines whether the rank order of the grammatical categories may have been influenced by the participants' (in)ability to process the semantic meaning of (some of) the test items belonging to each grammatical category.

Considering that the rank orders found for P1, P2 and P3 were highly similar (3.3.2) and, moreover, the number of *distractor* answers selected by P1, P2 and P3 over the entire GT were not significantly different either (3.1.3), this section will use the rank order that is found when calculating the mean score of the 72 participants – i.e. without classification of the participants into L2 contact groups. As can be seen in Table 9, this overall rank order was 1 NEG, 2 SVO, 3 GEN, 4 PLU, 5 POSS, 6 COP, 7 PROobj, 8 VERB and 9 PROsubj. Some deviations from the rank orders found in P1, P2 and P3 can be discerned in the five lowest-ranked categories. However these changes may be meaningless because of the statistical insignificance of the mean score differences (in P1, P2 and P3 and in the overall rank order).

An ANOVA showed a high level of significance; $F(8, 639) = 45.23$, $p < .001$. A Tukey HSD revealed that in order to be significant at the $p < .05$ level, the difference in mean score between 2 categories had to exceed 10.64; for significance at the $p < .01$ level, the difference had to be greater than 12.23. For the sake of completeness Table 9 records the significance of the inter-categorical mean score differences.

Table 9 *Overall Rank Order: Significance of Mean Score Differences*

	NEG *M* = 89.58 *SD* = 15.68	SVO *M* = 81.48 *SD* = 25.87	GEN *M* = 72.22 *SD* = 20.75	PLU *M* = 60.42 *SD* = 20.83	POSS *M* = 53.70 *SD* = 20.79	COP *M* = 48.84 *SD* = 17.32	PROobj *M* = 48.38 *SD* = 21.88	VERB *M* = 47.45 *SD* = 17.17	PROsubj *M* = 47.22 *SD* = 21.67
NEG *M* = 89.58 *SD* = 15.68									
SVO *M* = 81.48 *SD* = 25.87	*n/s*								
GEN *M* = 72.22 *SD* = 20.75	**p < .01**	*n/s*							
PLU *M* = 60.42 *SD* = 20.83	**p < .01**	**p < .01**	**p < .01**						
POSS *M* = 53.70 *SD* = 20.79	**p < .01**	**p < .01**	**p < .01**	*n/s*					
COP *M* = 48.84 *SD* = 17.32	**p < .01**	**p < .01**	**p < .01**	**p < .05**	n/s				
PROobj *M* = 48.38 *SD* = 21.88	**p < .01**	**p < .01**	**p < .01**	**p < .05**	*n/s*	*n/s*			
VERB *M* = 47.45 *SD* = 17.17	**p < .01**	**p < .01**	**p < .01**	**p < .05**	*n/s*	*n/s*	*n/s*		
PROsubj *M* = 47.22 *SD* = 21.67	**p < .01**	**p < .01**	**p < .01**	**p < .05**	*n/s*	*n/s*	*n/s*	*n/s*	

It has been explained before that when a participant selected a *distractor* it may be deduced that s/he did not understand the semantic content of the prompt. Replacing a prompt which yielded a *distractor* reply by a prompt which does not pose any semantic processing problems would result in either a *correct* response – if the grammatical category had been acquired – or an *error* reply – if the grammatical category had not been acquired. Consequently it cannot be said whether the grammatical category which was tested by means of the prompt had been acquired or not. This further means that counting the *distractor* replies as an incorrect answer, along with the *error* replies, may give a misleading account of the participants' level of receptive knowledge.

Table 10 below clearly shows that some categories contained more semantically problematic test items than other categories. In COP, PROobj and PROsubj, for example, 10% of the replies (in total) consisted of a distractor. In all other categories, the number of distractor replies remained below 5%.

Table 10 *Overall Rank Order: Proportion of Distractor Replies*

	NEG	SVO	GEN	PLU	POSS	COP	PRO obj	VERB	PRO subj
Correct (% out of total)	89.58	81.48	72.22	60.42	53.70	48.84	48.38	47.45	47.22
Error (% out of total)	7.87	12.96	22.92	37.50	43.06	40.28	39.58	49.31	41.20
Distractors (% out of total)	2.55	5,56	4.86	2.08	3.24	10.88	12.04	3.24	11.57
Incorrect (% out of total)	*10.42*	*18.52*	*27.78*	*39.58*	*46.30*	*51.16*	*51.62*	*52.55*	*52.78*

As has also been explained in sections 3.1.3 (and in section 2.5.3), it seems reasonable that the number of *distractor* replies which would change into a correct response if they were semantically understood stands in relation to the ratio of *correct* and *error* replies in the already unproblematic test items. Thus it is possible to gauge the effect of semantic processing problems on the rank order by discarding *distractor* responses and calculation the percentage of correct replies out of the total of semantically unproblematic replies.

The result of this calculation can be found in Table 11. The 'former rank' and 'former score' are those reported above, i.e. by considering distractors as incorrect responses (along with the *errors*). They are included in this Table to facilitate the comparison.

Table 11 *Rank Order Found by Excluding Distractors*

Rank Order	NEG	SVO	GEN	PLU	POSS	PRO obj	COP	PRO subj	VERB
Score	91.92	86.28	75.91	61.70	55.50	55.00	54.80	53.40	49.04
Former Rank Order	NEG	SVO	GEN	PLU	POSS	COP	PRO obj	VERB	PRO subj
Former Score	89.58	81.48	72.22	60.42	53.70	48.84	48.38	47.45	47.22

Two changes in rank order are evident in Table 11. PROobj and COP have switched positions, as have PROsubj and VERB. However, the difference between these two ranks were insignificant before, and are most likely still insignificant – given the small size of the mean score difference.

For other categories the differences in mean scores have changed somewhat. For example the difference between SVO and GEN was 9.26 in the 'former rank' but has increased to 10.37 in the new rank.

The results in Table 11 will not be examined in more detail (for example, through an ANOVA) because they are merely speculative, based on the assumption that the ratio of errors and correct responses within the semantically problematic responses would be similar to the ratio of errors and correct responses found in the semantically unproblematic responses. Thus it can merely be interpreted as a potential illustration of the role played by semantically difficult test items on the GT result. This role then appears not to be strikingly large. Even so it should be emphasized that a GT in which all tests items were semantically understood by the participants would provide more valid results and hence a more valid rank order.

3.3.4 Overall Rank Orders and the Role of Paired Test Items

As explained, the test items in GT/A and GT/B are paired, with one item always containing one 'alternative' of the grammatical category, and the other category containing an other 'alternative'. More specifically, in the categories PROobj, PROsubj and POSS the alternatives consist of the *masculine* and *feminine* 3rd person singular

pronoun. In PLU, the alternatives are *singular* versus *plural* – or in other words no plural marker versus a plural marker. In the categories COP and VERB the alternatives are the *singular* versus the *plural* 3rd person singular verb form – or in the case of VERB 3rd person 3SG *–s* versus no 3 SG *–s*. In the category NEG, the sentences in the test items are either affirmative (i.e. absence of negator *not*) or negated (i.e. containing the negator *not*). In the paired test items for SVO, the subject and object in the sentence are reversed by way of alternatives. Accordingly test items in this category cannot be classified as either one of the alternatives.

Importantly, in the categories PLU, NEG, GEN and VERB the test items which do not contain a morpheme may be considered easier to acquire than their alternatives on the basis that the former are less marked than the latter. As Ellis (2008) explains, the markedness of grammatical structures depends on several factors, such as frequency of occurrence (with less frequent structures being more marked), versatility (referring to the positions where it may occur and the number of inflections it may take, whereby a structure is more marked when it is less versatile) and whether or not morphemes are added (with more morphemes added to it resulting in a more marked structure). Accordingly, PLU: *plural*, NEG: *negation*, GEN: *marker present* and VERB: *singular* are more marked than their alternative PLU: *singular*, NEG: *affirmative,* GEN: *marker absent* and VERB: *plural* and consequently may be easier to acquire.

In the categories POSS, PROobj and PROsubj, conversely, there is no reason to assume that either one of the alternatives (masculine or feminine) is easier to acquire, since the literature on markedness does not suggest that masculine is less marked than feminine or vice versa.

In COP, finally, the alternative *is* and *are* are both marked where morphology are concerned, and hence may be equally difficult to acquire.

Table 12 presents the nine grammatical items, their mean score and the mean score attained on each of the alternatives, along with the statistical significance as revealed by paired, two-tailed *t*-tests.

Table 12 *Paired Test Items: Mean Scores Attained on Alternative Test Items*

Grammatical Categories	Alternatives	Overall Mean	Alternatives Mean	*t*(71)	*p*
NEG	Affirmative Negation	89.59	91.67 87.50	1.45	(0.15)
SVO	/ /	81.48	/ /	/	/
GEN	No GEN GEN	72.22	78.24 66.20	2.93	0.01
PLU	Sg Plu	60.42	79.63 41.20	6.4	<0.001
POSS	M F	53.71	56.48 50.93	1.04	(0.3)
COP	Sg Pl	48.84	62.96 34.72	4.30	<0.001
PROobj	M F	48.38	42.59 54.17	2.3	0.02
VERB	Sg Pl	47.45	61.57 33.33	4.47	<0.001
PROsubj	M F	47.22	45.83 48.61	0.47	(0.6)

In line with what was expected, in the category GEN and more strikingly so in the category PLU participants attained a significantly higher score on the unmarked test items than on the marked test items. Furthermore, it may be observed that the score attained on the morphologically marked test items in PLU was lower than the overall mean scores and alternative category mean scores (except for COP Plural and VERB Plural) attained in the categories below PLU. Consequently a question that arises is whether the score and by extension the rank position of PLU is truly representative of the participants' knowledge of the category. Averaging the marked and unmarked category of course compensates for the high score attained on the unmarked category. On the other hand it should also be taken into account that in POSS, PROobj and PROsubj both alternatives are marked, or at least more marked and consequently more difficult than PLU: *singular*.

With regard to the category GEN, it was speculated in section 2.2.2 that, when processing test items containing a genitive -*'s*, participants might rely on the final word in the sentence (e.g. *boy* vs. boy's *dog*) to correctly answer the question. While this clearly did not suffice for attaining an equally high score on the test items with a genitive –'s as on the test items without this morpheme, it is not unthinkable that at least some of the correct replies for this score are not the result of grammatical mastery of the feature GEN. Hence the receptive knowledge of the category GEN might also not be as high as the overall score suggests. Whether the rank order is invalid cannot be said.

In both COP and VERB the scores attained on 3rd person singular were higher than the scores on 3rd person plural, reaching high levels of significance. In the case of COP, this is remarkable because both categories are equally marked. In the case of VERB the result is even more surprising because the plural form is unmarked, whereas the singular form is marked.

A possible explanation is that certain participants were unable to distinguish 3rd person singular and plural but grasped the meaning of the subject and the verb and with this information opted for the picture representing singular. The zero-plural of the sentence subject may have contributed to this, in that participants interpreted the prompt as containing a singular subject and verb. Support for the latter comes from a few participants who mastered the category PLU but asked, before pointing to the picture containing a singular subject, whether the subject was '*sheep'* or '*sheeps'*. Again, however, no conclusive evidence to support this hypothesis can be found in this study.

Lastly, in the category NEG, the scores on the marked test items were found to be slightly lower than the scores on the unmarked test items, but not significantly so. Consequently there is no reason to question the score and rank of this category. As expected, the score of the two alternatives in the categories POSS and PROsubj were not significantly different either. Defying all explanation, PROobj: *feminine* appeared to be better acquired than PROobj*: masculine*.

In short, the above discussion has indicated that the nature of the paired test items may be such that care must be taken in interpreting some of the scores. In case of significant differences between two alternative, averaging the two scores compensates for both high and low scores. Because in some categories one alternative was distinctively less marked and hence less difficult (PLU, GEN) than the other alternative, while in other categories both alternatives were equally marked (PROobj, PROsubj, and POSS) and hence equally difficult, the possibility that the scores on the former categories may not be truly representative of the participants' knowledge of the category cannot be altogether rejected.

3.3.5 Summary

The inter-categorical differences in P1, P2 and P3 showed very little statistical significance. The six categories for which the lowest level of receptive grammar

knowledge was observed are PLU, POSS, COP, PROobj, PROsubj and VERB. Some changes can be observed in the rank order of these 6 items when comparing P1, P2 and P3, but in none of the three groups the differences in mean scores were statistically significant – with the exception of a significant difference in mean scores between PLU (rank 4) and PROobj (rank 9) in P2. The order of the three highest-ranked items (NEG, SVO and GEN) was constant over the three L2 contact groups. Furthermore, disregarding some minor changes in the level of significance ($p < .05$) and ($p < .01$) as well as some shifts between significance and insignificance in P1, P2 and P3, the scores attained in each of these categories was significantly different from the scores attained in the eight other grammatical categories, except for the category – or in some cases two or three categories – ranked just above or below.

Given the high correlation between the rank orders found for P1, P2 and P3, a fourth rank order was made, based on the scores of all participants (P1, P2 and P3 together). This rank order can be found in Table 13.

Table 13 *Rank Order of Nine Grammatical Categories*

Rank	Grammatical Features
1	NEG
	SVO
2	GEN
3	PLU
4	POSS
	COP
	PROobj
	VERB
	PROsubj

SVO and NEG are both given the first rank on the basis of the statistically insignificant difference between the two scores. More research is needed, however, to determine whether they are interchangeable in rank. In keeping with the significantly different GT scores, GEN and PLU are given the 2nd and 3rd rank respectively. The remaining categories are ranked together, for a number of reasons. First, the difference in scores was not statistically significant. Secondly, while the order in Table 12 is in accordance with the mean scores across the different L2 contact groups, the orders found for P1, P2 and P3 as well as the order found in the implicational scales diverge from this rank.

Finally, PLU had a significantly higher score than COP, PROobj, VERB, and PROsubj. The difference with POSS, however, was not statistically significant. Notwithstanding this finding, PLU and POSS are assigned a different rank, for two reasons. First, the score of POSS was not statistically different from the four categories ranked below it. In addition, in the rank orders based on the scores of the separate L2 contact groups, the rank of POSS varies. Hence, it seemed more acceptable to place POSS in the fourth rank, along with COP, PROobj, VERB, PROsubj than to separate it from these four categories and rank it alongside PLU.

To ascertain whether semantic comprehension problems did not influence the rank order of the grammatical categories, the number of *distractor* answers for each of the categories was studied. It was concluded that no *significant* changes would occur if the semantically difficult prompts were to be replaced by easier ones. An examination of the paired test items, further, revealed that the scores of certain categories may not have been in accordance with the participants' receptive knowledge of these categories. Notably the scores for PLU and possibly also GEN, COP and VERB may give too positive an impression of the participants' receptive knowledge of these categories. Whether the rank order is affected by this cannot be said, however.

To conclude, it has been explained in the literature study (Chapter I, section 2.1.2) that rank orders based on accuracy of usage (as is the case in cross-sectional studies) can only be interpreted as an acquisition order if it is assumed that a high accuracy score points to an early acquisition, and a lower score to a later acquisition. Thus, the order in Table 13 - which is based not on accuracy of usage but on GT scores – should not be interpreted too readily as also reflecting the acquisition order.

3.4 Developmental Patterns in the Acquisition of Receptive Grammar Knowledge

The previous sections have rank ordered the grammatical categories on the basis of the mean scores. Notwithstanding that a high degree of correlation between the rank order in P1, P2 and P3 was attested, it would be premature to conclude that these rank order also embody an acquisition order, not in the least because the data are cross-sectional and because few mean score differences between the categories were significant. In addition, rank orders have been criticised for obscuring individual variation, thus suggesting systematicity where this may not be justified (Ellis, 2008).

With this in mind, this last section seeks to determine whether it is possible to speak of systematicity in the acquisition of receptive grammar knowledge by means of *implicational scaling*. Implicational scaling is considered a more reliable method for this purpose, although it must be conceded that it is not entirely uncontested either (Hatch & Farhady, 1982). Furthermore, implicational scaling has not yet been used with receptive grammar tests such as the one used in this study.

Two implicational scales have been made in this section. In the first one, a grammatical item is considered *acquired* if the participant has received an average score of 90% or higher on the category in question. As explained in the methodology, however, only scores of 100% fulfil this requirement: 5 correct answers out of 6 test items converts to a score of 83.33 percent; thus only 6 out of 6 or 100 % is higher than 90%. Therefore a second scale was made, this time taking 80% as the minimum score for an item to be considered acquired. Because 80% is a somewhat less common criterion for acquisition, the 80% scale will not altogether replace the 90% scale. Rather, as said, the two scales are presented alongside each other.

The scales have been included at the end of this section. Table 14 contains the 90% scale, Table 15 the 80% scale. The step-like line running through each of the scales separates the features that were acquired (represented by 1 in the scale) from the features that were not acquired (represented by 0). Deviations from the ideal pattern – i.e. scale only 1s to the left of the line, only 0s to the right of the line – are placed between brackets.

In neither of the scales any participants can be found who had acquired all nine grammatical categories. In the 90% scale one participant had acquired five items; all others had acquired four items or less. In the 80% scale one participant had acquired seven items, the others six items or less. Thus, neither of the scales is complete.

A common result of changing the criterion for acquisition (Hatch & Farhady, 1982) is that the rank order of the grammatical items changes. In this study, however, the rank orders found in the 80% scale and in the 90% scale correspond. By contrast the rank order in these scales does deviate from the rank order(s) based on the GT score. Divergences only occur in the four lowest ranks, however, which have been shown not to have significantly different GT scores.

In both scales, a number of deviations from the ideal pattern can be found, situated across all grammatical categories. Despite these deviations, in both Table 14 (90% scale) and Table 15 (80% scale) the data were found to be scalable.

In the 90% scale, C_{rep} (or *IR*) was calculated as .94, meaning that 94 times out of 100 the expected pattern of 0s and 1s was fulfilled. It is generally agreed that C_{rep} should be higher than .90. Although this is sometimes accepted as sufficient evidence that the data are scaled, as a further test the coefficient of scalability can be calculated. MM_{rep} was found to be .17, which gives a coefficient of scalability of .93. This is well above the minimum for a set of data to be scalable, set at .60.

In the 80% scale, C_{rep} was found to be only just .90, making it borderline acceptable. MM_{rep} was .30, giving an acceptable coefficient of scalability of .86.

In short, to the extent that it is possible to draw conclusion on the basis of an incomplete implicational scale containing grammatical categories with few significant rank orders, this section indicates that there is considerable systematicity in the receptive acquisition of the nine grammatical categories in this section.

Table 14 *Implicational Scale: 1 = more than 90% correct; 0 = less than 90% correct*

ID	NEG	SVO	GEN	PLU	POSS	PROobj	PROsubj	COP	VERB
ID62	1	1	1	(0)	1	0	(1)	0	0
ID57	(0)	1	1	1	(1)	0	0	0	0
ID51	1	1	1	0	0	0	0	0	0
ID30	1	1	1	0	0	0	0	0	0
ID42	1	1	1	0	0	0	0	0	0
ID43	1	1	1	0	0	0	0	0	0
ID54	1	1	(0)	0	0	(1)	0	0	0
ID66	1	1	(0)	(1)	0	0	0	0	0
ID71	1	1	1	0	0	0	0	0	0
ID29	1	1	1	0	0	0	0	0	0
ID59	1	1	1	0	0	0	0	0	0
ID39	1	1	1	0	0	0	0	0	0
ID72	1	1	(0)	(1)	0	0	0	0	0
ID60	1	1	1	0	0	0	0	0	0
ID9	1	1	1	0	0	0	0	0	0
ID21	1	1	1	0	0	0	0	0	0
ID61	(0)	1	0	(1)	0	0	0	0	0
ID65	1	1	0	0	0	0	0	0	0
ID7	1	1	0	0	0	0	0	0	0
ID34	1	1	0	0	0	0	0	0	0
ID1	1	1	0	0	0	0	0	0	0
ID10	1	1	0	0	0	0	0	0	0
ID24	(0)	(0)	(1)	(1)	0	0	0	0	0
ID31	(0)	1	0	(1)	0	0	0	0	0
ID70	1	1	0	0	0	0	0	0	0
ID38	1	1	0	0	0	0	0	0	0
ID46	1	1	0	0	0	0	0	0	0
ID19	1	1	0	0	0	0	0	0	0
ID48	1	1	0	0	0	0	0	0	0
ID28	1	1	0	0	0	0	0	0	0
ID53	1	1	0	0	0	0	0	0	0
ID2	(0)	(1)	0	0	0	0	0	0	0
ID8	1	0	0	0	0	0	0	0	0
ID67	1	0	0	0	0	0	0	0	0
ID27	1	0	0	0	0	0	0	0	0
ID63	1	0	0	0	0	0	0	0	0
ID4	1	0	0	0	0	0	0	0	0
ID15	1	0	0	0	0	0	0	0	0
ID17	(0)	0	0	0	(1)	0	0	0	0
ID50	(0)	0	(1)	0	0	0	0	0	0
ID69	(0)	(1)	0	0	0	0	0	0	0
ID6	1	0	0	0	0	0	0	0	0
ID41	1	0	0	0	0	0	0	0	0
ID45	1	0	0	0	0	0	0	0	0
ID36	1	0	0	0	0	0	0	0	0
ID44	(0)	(1)	0	0	0	0	0	0	0
ID55	(0)	(1)	0	0	0	0	0	0	0
ID68	(0)	(1)	0	0	0	0	0	0	0
ID3	1	0	0	0	0	0	0	0	0
ID5	(0)	(1)	0	0	0	0	0	0	0
ID12	(0)	(1)	0	0	0	0	0	0	0
ID58	1	0	0	0	0	0	0	0	0
ID52	1	0	0	0	0	0	0	0	0
ID13	1	0	0	0	0	0	0	0	0
ID26	(0)	(1)	0	0	0	0	0	0	0
ID23	(0)	0	0	0	0	(1)	0	0	0
ID49	1	0	0	0	0	0	0	0	0
ID37	1	0	0	0	0	0	0	0	0
ID32	1	0	0	0	0	0	0	0	0
ID64	0	0	0	0	0	0	0	0	0
ID47	0	0	0	0	0	0	0	0	0
ID14	0	0	0	0	0	0	0	0	0
ID40	0	0	0	0	0	0	0	0	0
ID16	0	0	0	0	0	0	0	0	0
ID56	0	0	0	0	0	0	0	0	0
ID22	0	0	0	0	0	0	0	0	0
ID35	0	0	0	0	0	0	0	0	0
ID25	0	0	0	0	0	0	0	0	0
ID33	0	0	0	0	0	0	0	0	0
ID11	0	0	0	0	0	0	0	0	0
ID18	0	0	0	0	0	0	0	0	0
ID20	0	0	0	0	0	0	0	0	0

Table 15 *Implicational Scale: 1 = more than 80% correct; 0 = less than 80% correct*

ID	NEG	SVO	GEN	PLU	POSS	PROobj	PROsubj	COP	VERB
ID62	1	1	1	1	1	(0)	1	(1)	0
ID61	1	1	(0)	1	1	1	(1)	0	0
ID57	1	1	1	1	1	1	0	0	0
ID65	1	1	1	(0)	(0)	(1)	(1)	0	0
ID66	1	1	1	1	1	0	0	0	0
ID29	1	1	1	1	(0)	0	0	(1)	0
ID39	1	1	1	1	(0)	(1)	0	0	0
ID51	1	1	1	(0)	0	0	(1)	0	0
ID7	1	1	(0)	1	0	0	(1)	0	0
ID30	1	1	1	1	0	0	0	0	0
ID42	1	1	1	1	0	0	0	0	0
ID43	1	1	1	1	0	0	0	0	0
ID54	1	1	1	(0)	0	(1)	0	0	0
ID1	1	1	1	1	0	0	0	0	0
ID72	1	1	(0)	1	(1)	0	0	0	0
ID27	1	1	1	(0)	(1)	0	0	0	0
ID9	1	1	1	1	0	0	0	0	0
ID31	1	1	1	1	0	0	0	0	0
ID63	1	1	1	(0)	0	0	(1)	0	0
ID24	(0)	1	1	1	0	0	0	0	0
ID71	1	1	1	0	0	0	0	0	0
ID59	1	1	1	0	0	0	0	0	0
ID2	1	1	1	0	0	0	0	0	0
ID8	1	1	(0)	0	0	0	0	(1)	0
ID67	1	1	1	0	0	0	0	0	0
ID60	1	1	1	0	0	0	0	0	0
ID70	1	1	(0)	0	(1)	0	0	0	0
ID4	1	(0)	1	0	0	(1)	0	0	0
ID15	1	1	(0)	0	0	1	0	0	0
ID17	1	(0)	1	0	(1)	0	0	0	0
ID19	1	1	1	0	0	0	0	0	0
ID21	1	1	1	0	0	0	0	0	0
ID45	1	1	1	0	0	0	0	0	0
ID36	1	1	1	0	0	0	0	0	0
ID28	1	1	(0)	0	0	0	0	0	(1)
ID53	1	1	1	0	0	0	0	0	0
ID37	1	1	1	0	0	0	0	0	0
ID50	(0)	1	(1)	0	0	0	0	0	0
ID69	(0)	1	0	0	0	0	0	(1)	0
ID47	(0)	1	0	0	0	0	(1)	0	0
ID68	(0)	1	(1)	0	0	0	0	0	0
ID26	(0)	1	0	(1)	0	0	0	0	0
ID34	1	1	0	0	0	0	0	0	0
ID10	1	1	0	0	0	0	0	0	0
ID38	1	1	0	0	0	0	0	0	0
ID46	1	1	0	0	0	0	0	0	0
ID6	1	(0)	0	0	(1)	0	0	0	0
ID48	1	1	0	0	0	0	0	0	0
ID14	1	(0)	0	0	(1)	0	0	0	0
ID44	1	1	0	0	0	0	0	0	0
ID55	1	1	0	0	0	0	0	0	0
ID3	1	(0)	0	(1)	0	0	0	0	0
ID5	1	1	0	0	0	0	0	0	0
ID12	1	1	0	0	0	0	0	0	0
ID40	1	(0)	0	0	0	0	0	0	(1)
ID58	1	1	0	0	0	0	0	0	0
ID13	1	1	0	0	0	0	0	0	0
ID49	1	1	0	0	0	0	0	0	0
ID64	(0)	0	(1)	0	0	0	0	0	0
ID16	(0)	0	(1)	0	0	0	0	0	0
ID23	(0)	0	0	0	0	(1)	0	0	0
ID41	1	0	0	0	0	0	0	0	0
ID52	1	0	0	0	0	0	0	0	0
ID56	1	0	0	0	0	0	0	0	0
ID35	1	0	0	0	0	0	0	0	0
ID25	1	0	0	0	0	0	0	0	0
ID32	1	0	0	0	0	0	0	0	0
ID11	1	0	0	0	0	0	0	0	0
ID18	1	0	0	0	0	0	0	0	0
ID22	0	0	0	0	0	0	0	0	0
ID33	0	0	0	0	0	0	0	0	0
ID20	0	0	0	0	0	0	0	0	0

4 Discussion

4.1 Rate of Receptive Second Language Grammar Acquisition

In seeking to answer the first research question concerning rate of receptive grammar acquisition, this study has examined the overall GT results, so as to ascertain whether the participants' receptive grammar knowledge expands as a result of L2 contact.

In summary, the data analysis has shown that after 18 months of L2 contact (L2 contact group P1), the participants' reached a mean score of 58% on the GT. A comparison of the L2 contact groups P1, P2 and P3 (with 18, 29 and 41 months of L2 contact respectively) showed that the progress made by the participants over the next two years was much more modest. In P3, the participants' mean score of 66% was significantly higher. In the intervening phase P2 the mean score was 61%, which, in addition to constituting only a minor increase in comparison to P1, was found to be significantly different neither from P1 nor from P3.

Furthermore, it was found that the participants in P2 and P3 had a similar score on GT/A and GT/B (the two parts of the test that were performed on different days) whereas in P1 the scores on GT/A (56%) and GT/B (60%) were less consistent. This might suggest that the participants receptive grammar knowledge was less stable in P1 than in P2 and P3, in that the former L2 contact group's ability to correctly answer a question varies more from one occasion to another and is more easily affected by factors such as anxiety and familiarity with the test.

The data for the nine grammatical categories were also examined in more detail. It was found that few of the increases and decreases in scores attested when comparing P1, P2 and P3 were significant. SVO showed a significant increase in score between P1 and P3 while POSS was marked by a significant increase between P2 and P3. As for the other results, the variegated rises and falls in scores suggested that each category is acquired at its own rate, with some categories showing a steady progress and others a more regular behaviour, which in some cases is even marked by U-shaped behaviour. However, owing to a lack of statistical significance no valid conclusions can be drawn.

An important factor in discussing the results of this study is age of acquisition. The participants' age of onset in acquiring English was around 5 to 6 years, which means that they had already passed the cutoff point (usually situated at around 4 years of age;

Ortega, 2009b) at which newly learned languages are acquired as an L2 rather than as an additional L1. They can still be considered very young L2 learners however.

Age of acquisition is a much-debated topic in SLA. A common folk wisdom goes that young children 'absorb languages like sponges', and in the scholarly field of SLA too it has been a long-held belief that young children learn new languages with greater ease and speed than adults. The age at which L2 learners are considered 'adults' varies, but often-found break points are 'seven years of age' or 'around the beginning of puberty' (Ellis, 2008; Gass & Selinker, 2008).

The explanations that have been given for this ease of acquisition are too numerous to discuss in this study, ranging from cognitive over neurological to affectional factors. More important is that many scholars now challenge this position. In particular, there is a growing consensus that young L2 learners may have the advantage of ultimate attainment, in that they reach native-like proficiency more often than late learners, but late(r) L2 learners are believed by many to have the advantage of rate. In the initial years of L2 acquisition especially, they have been found to acquire the new language faster than young learners. Because of various factors (e.g. more as well as more intensive L2 contact among young learners, to name just one example), young learners eventually catch up with and usually also surpass older learners (Baker, 2006; Ellis, 2008).

In addition to the opposition between rate and ultimate attainment, scholars also often stress the differences between *how* and *where* young and older L2 learners acquire an L2 most efficiently. For example, there is widespread agreement that older learners may require formal L2 (grammar) instruction in order to reach high levels of ultimate attainment. Young learners, by contrast, do not have the cognitive maturity to procees formal instruction and hence will fare better in naturalistic learning environments (Baker, 2006; Ellis, 2008).

The score of nearly 60% attained by participants in P1, after 18 months of L2 contact, might suggest a fairly fast rate of acquisition over the first year and a half of L2 acquisition. However, results would have to be compared with test results gathered at the onset of L2 acquisition and after, for example, 6 months of L2 contact for more valid conclusions. Furthermore, a comparison of the scores attained by P2 and P3 showed a small increase of only 3 to 5 % over the course of one year, giving a modest

increase in score of 8 % over the course of two years. This could be taken to suggest a considerably slower rate of acquisition – although of course it is debatable how large an increase in scores is required in order to speak of a fast rate of acquisition. The scores attained on the distinct grammatical categories – which revealed a very irregular development marked by minor (and for the most part insignificant) rises *and* falls in scores – further support this slow rate of acquisition.

Before extrapolating the results of this study to the acquisition of receptive grammar acquisition among young L2 learners in general, a number of factors must be considered which may account for the slow rate of acquisition found in this study. A first factor that will be discussed is concerned with the GT that has been used. Other potential explanations are related to the data that have been obtained through the GT, the design of the study, and the specific setting in which the participants in this study were acquiring the L2.

With regard to the GT, it warrants repeating that only 9 grammatical categories were included in this study, some of which were furthermore very closely related (notably PROobj, PROsub and POSS). More or different grammatical categories may give a different result.

Concerning the data that were analysed, it should be noted that individual variation among the participants in the three L2 contact groups was high. As demonstrated in the results, some participants in P1 attained a score that was higher than participants in P3. Given that standard variation is an important component in calculating statistical significance (Hatch & Farhady, 1982), the lack of statistical significance between P1 and P2 and between P2 and P3 might partially result from this individual variation. Another factor is that the sample of participants in this study may have been too small. However, statistically significant or not, an increase in score with 3 to 5 % over the course of a year remains evidence more in favour of a slow rate of acquisition than in favour of a fast rate.

A fourth factor that must be considered is the cross-sectional design of the study, which in itself calls for caution, in combination with the fact that the immersion programme in which the participants were enrolled had been implemented in the Ecole Cheval Bayard only a few years earlier. The participants in P3 were the school's first generation of immersion pupils. Along with the staff problems in the school and shortage of materials

– as observed during the participant observation mentioned earlier –, the cross-sectional design and recent implementation of the immersion school may go some way in accounting for the modest differences in scores between the different age groups.

Finally, and most importantly perhaps, there is evidence that whereas route of acquisition may be impervious to learning context, rate of acquisition is not (see Chapter I, section 4). Consequently the immersion setting itself can be considered in important factor. Immersion education falls under Ellis' (2008) definition of 'instructed language acquisition' as 'a language acquisition that takes place as a result of attempts to teach the L2 – either directly through formal instruction or indirectly by setting up the conditions that promote natural acquisition in the classroom' (p. 966). Despite this classification of immersion education as an instructed learning context, grammar acquisition in immersion education is very much an *implicit* process: no, or at most very little, explicit attention is paid to grammar. Rather, the focus in immersion classrooms is on 'meaningful communication', and the acquisition of L2 aspects such as grammar and vocabulary is expected to occur as a side effect (Baker, 2006; Baker & Prys Jones, 1998).

Research in immersion schools has shown that, whereas the pupils reach high levels of comprehension skills in the L2 and can express themselves fluently, very frequently a fossilisation of the pupils' grammar knowledge is attested. It is argued that this is the result of the 'focus-on-meaning' approach (in the form of explicit grammar instruction), in combination with the limited L2 output by the pupils (Cummins, 2000; Genesee, 2004; Swain, 1995). As mentioned in chapter II (section), however, in order for grammar acquisition to occur, semantic comprehension of the message may not suffice. Rather, input must be processed at an analytic (i.e. grammatical) level as well. As a result of the focus-on-meaning approach, however, the pupils' processing of the input is very much focused on the comprehension of the general meaning rather than on grammatical analysis. As has also been mentioned in chapter II, furthermore, many scholars agree that not only input but also output is a necessary requirement for L2 learning to take place, a requirement which may not be sufficiently fulfilled in immersion education.

Note that the fossilisation of grammar skills attested in immersion education relates to language production only. Furthermore studies on grammar acquisition in immersion

education have focused on the eventual outcome of immersion programmes, and hence they do not provide evidence as to whether the rate at which grammar knowledge is acquired is also affected. It appears reasonable to assume, however, that this will be the case. The results from this study further support this.

As a side note, the small progress made by the participants in this study need not mean that immersion education is not an efficient method for learning languages. This study has only provided data with regard to early years of acquisition. Consequently it would be premature to draw conclusions with regard to the eventual outcome of the programme. The receptive grammar acquisition process may for example accelerate in later years.

In conclusion, with regard to rate of acquisition both the overall GT results and the more detailed analysis of the nine grammatical categories constitute evidence that the acquisition of receptive grammar knowledge takes place at a relatively slow pace, at least among young L2 learners in immersion education. To extrapolate the findings from the acquisition of receptive grammar knowledge by young learners in immersion settings to receptive grammar acquisition more generally, however, more research in different L2 learning settings and with learners of different ages is required. Furthermore certain shortcomings of the GT used and of the data, such as semantic difficulty and too small participants samples, would have to be remedied.

4.2 Route of Receptive Second Language Grammar Acquisition

4.2.1 Rank Orders and Scalability

The second, primary research question of this study was whether developmental stages can be discerned in the acquisition of receptive grammar knowledge.

The analysis – discussed above – of the progress made in the nine grammatical categories, was meant to give some first insights into possible patterns. While conclusions concerning the progress made in the categories were impossible due to problems of statistical insignificance, the graph presented in section 3.2.1 (Figure 3) also formed a good illustration of the rank order of the grammatical categories based on the scores attained on each of the categories by participants in P1, P2 and P3.

A detailed analysis of these rank order showed a high degree of correlation between the rank orders in P1, P2 and P3, so that is was deemed acceptable to make one rank order based on the scores of all 72 participants, without classification into L2 contact groups.

The rank order for the 9 grammatical categories was found to be as follows:

Rank	Grammatical Features
1	NEG
	SVO
2	GEN
3	PLU
4	POSS
	COP
	PROobj
	VERB
	PROsubj

A statistical analysis of this rank order led to the conclusion that a more detailed rank order (with each category being assigned a different rank) would be groundless. Moreover, an evaluation of the effect of semantic comprehension and an examination of the scores attained on the two 'alternatives' within the paired test items of each grammatical category warned for the possibility that even the above rank order may not be altogether correct.

This rank order is based on GT scores in a cross-sectional study and can only be interpreted as an acquisitional order if it is assumed that the items that are more difficult have been acquired last, while the items that are easy have been acquired first. Furthermore individual variation is obscured in a rank order based on mean scores. Therefore implicational scaling, a more acceptable way of assessing whether this order is indeed an acquisition order and whether it is possible to speak of systematicity in the acquisition of these grammatical items, has been used.

Both an implicational scale made with an 80% criterion of acquisition and a scale made with a 90% criterion of acquisition showed that the data were scalable, thus suggesting that the above rank order is also an implicational acquisition order. Owing to the incompleteness of the scale, however, this result must also be interpreted with some care.

4.2.2 Explaining Systematicity

As this study gives a tentative but nonetheless positive answer to the question whether there is systematicity in the acquisition of L2 acquisition, a further question is concerned with how this acquisition order relates to language production.

As explained in Chapter I (section 2.1.1), the acquisition order found in morpheme studies was 1) -ing, *plural*, and *copula*, 2) auxiliary and article, 3) irregular past, 4) regular past, *3rd person singular and possessive –s.* Variations may occur within these four ranks. Of the morphemes in this productive morpheme acquisition rank, four morphemes were also (partly) included in this study, namely plural (PLU), copula (COP), 3rd person singular (VERB), and possessive –s (POSS). The acquisition order found in morpheme studies thus clearly does not correspond with the rank order found in this study of receptive grammar acquisition (see table on page 88 above).

As a result of this divergent rank order, it is also unlikely that the explanations offered for morpheme acquisition orders found in productive grammar acquisition – such as frequency of occurrence – can account for the acquisition order found in this study.

Markedness, sometimes posited as offering a (partial) explanation for acquisition orders, has already been discussed (this chapter, section 3.3.4). It was found that grammatical features in which both alternatives in the paired test items were equally marked (POSS, PROobj, PROsubj, COP) were ranked lower than grammatical features in which one alternative was distinctly less marked (PLU, NEG, GEN). It may be added here that SVO, an unmarked word order, also attained a low score. Thus markedness plays some role in the rank order found in this test. However the degree of markedness of the nine distinct grammatical categories relative to one another would have to be determined in order to define the role of markedness with more certainty - a task which is hardly possible when using a GT in which one grammatical category is tested by means of marked *and* less marked variants of a grammatical category.

Also difficult to assess is the validity of Processability Theory (Chapter I, section 3.1.4), because not all processing procedures are represented in the GT.

When it comes to language production, PT clearly explains which processing procedures are required. For example, the formation of PROobj and PROsubj require the category procedure – no exchange of information between different constituents is required for their processing. Similarly, for plural formation (within one constituent),

the category procedure suffices. However, number agreement requires the phrasal procedure. Similarly, verb formation and formation of the possessive determiner require the category procedure while subject-verb agreement and gender agreement between the possessive determiner and the possessor require the S-procedure. In production, further, NEG would require the S-procedure. Finally, word orders other than SVO also require the S-procedure, but SVO itself, as the canonical word order, can typically be formed before this procedure is acquired.

In the ELIAS GT, category procedure suffices for most of the test items. In the category PLU the test items consist of only 1 lexical item. Consequently no other procedures are needed. In POSS test items consists of a phrasal structure (e.g. his cat), but only the category procedure would be required to recognize *his* and *her*. Likewise, PROsubj, PROobj and GEN also require only the category procedure. As these categories are embedded in longer clauses, other processing procedures would be at work, but technically speaking these would not be required to process the pronouns and the genitive – 's.

COP and VERB too are embedded in longer clauses, but again both categories can be processed by the category procedure only. Although in the original ELIAS GT the categories are called AGRc and AGRv respectively – with *AGR* standing for agreement, and *c* and *v* for copula and (full) verb respectively – there is no actual need for verb agreement in receptively processing these categories. Rather receptive knowledge of the verbal morphology of full verbs and copular verbs suffices. Since the sentence subjects have a zero plural (*sheep*, *fish*, *deer*) looking for the number of the subject would not help participants.

Thus far, then, the results of the grammar test shed no light on whether or not PT is applies to receptive grammar processing too. Furthermore, for the remaining two categories, SVO and NEG, it is much less clear which processing procedures would be required for receptively processing the categories. The high rank order of SVO and NEG could contradict the applicability of PT, as PT claims that the S-procedure required for processing these structures cannot be acquired before the category procedure. However, SVO might also have a high rank in the acquisition order because it is the canonical word order. Similarly, a correct answer of the negative sentences in

the GT may simply depend on the recognition of the negator *not*, which is inserted in the otherwise canonical SVO word order.

4.2.3 The Role of the First Language

L1 transfer is considered an important factor in L2 acquisition, both in language production and in language comprehension (Berthele et al., in press; Ellis, 2008). In their study of receptive grammar acquisition, Berthele et al. (in press) therefore argue that studies of grammar acquisition should evaluate the role played by the L1, so as to ascertain to what extent developmental patterns are universal or are influenced by the L1.

Section 2.2.2 has examined how the English grammatical categories relate to the pupils' L1. With regard to SVO, it was concluded that similarity to the L1 might facilitate acquisition. Similarly NEG may be an easy category because the position of the negator *not* in the SVO word order is similar. In the category PLU, by contrast, the fact that plural –s is not pronounced in the L1 French might, through negative transfer, cause a low score. In the category POSS, negative transfer is possible because of the difference with regard to how gender is determined. For other categories, neither negative nor positive transfer appears particularly likely.

In the rank order, SVO and NEG take up a high rank, which might point to L1 positive transfer. However PLU and POSS were both ranked higher than several other categories, suggesting that negative transfer may not be at work in receptive grammar acquisition.

For more insight into the role of transfer, a wider scope including a comparison of participants of various L1 backgrounds would be interesting. Such a comparison was not possible with the participants in this study because, although several participants were bilingual, all of them had French as a (in most cases dominant) L1. Furthermore each of the additional L1s was not represented by a large enough sample of participants.

5 Prospects for Future Research

This section has presented an exploratory investigation into receptive L2 grammar acquisition, an aspect of grammar acquisition which has been largely neglected in SLA. Contrary to language production, where years of research have produced an abundance

of research methodologies and test materials and have given insight into which types of investigations produce more reliable results than others, as yet it is far less clear how how receptive L2 grammar acquisition can best be investigated. It should be no surprise therefore that the results in this study are far from conclusive.

On the basis of problems and difficulties encountered in the course of this investigation, a few findings and conclusions can be formulated which might contribute to future research into receptive L2 grammar acquisition.

First, it may also be repeated here that, although research into grammar acquisition has often extrapolated findings from productive grammar acquisition studies, this extrapolation may be ill founded, in that it is based on the assumption that the acquisition of receptive knowledge not only always precedes the acquisition of productive skills, but is also marked by similar developmental patterns. Empirical investigations, anchored in but also critical of theories of L2 (grammar) acquisition, are needed to either support or dismiss this traditional view on language reception.

Potentially the most problematic factor encountered in this study was semantic comprehension. If grammatical features are embedded into longer phrases or clauses consisting of words or phrases which are not understood by the participants, the aim of the test – namely to show whether or not the grammatical feature is acquired – is defeated. If semantic processing fails, in particular with participants as young as the ones in this study and with such a limited knowledge of the L2, it is likely that the grammatical categories will not be processed – regardless of whether they were acquired or not. Consequently semantic comprehension will have to be controlled for in any receptive grammar test. Including distractors, as in this study, is only a partial remedy: while it shows where semantic processing was problematic, it does not allow for conclusions concerning the extent to which results have been affected. Adaption of the test items to the level of the participants' level of L2 proficiency may be a better solution – though, granted, perhaps not always feasible.

In many other respects, research into receptive L2 grammar acquisition resembles research concerned with productive L2 grammar acquisition, and as such will have to face the same challenges as has been and still are faced by researchers in this subfield of SLA. Before being able to generalize results, for example, research will have to be conducted using participants from different L1 backgrounds, of different ages and who

are learning the L2 in different settings. In addition different types of receptive grammar test need to be used and different aspects of grammar acquisition will have to be investigated in order to gain a truly comprehensive understanding of receptive L2 grammar acquisition. Furthermore, studies with not just cross-sectional but also longitudinal designs will have to be conducted. Finally, the test used in this study was not designed to test particular explanatory theories or hypothesis, and hence testing Processability Theory proved to be difficult. Thus, here, too, lies a challenge for future research.

6 Conclusion

This chapter has investigated the receptive grammar knowledge of 72 participants enrolled in the Ecole Cheval Bayard, an English immersion school. Participants were selected on the basis of the amount of L2 contact they had had at the time of testing, and could accordingly be classified into one of three cross-sectional L2 contact groups: 18 months of L2 contact (n = 28), 29 months of L2 contact (n = 23) or 41 months of l2 contact (n = 21).

Receptive grammar knowledge was assessed by means of a multiple choice test in which participants had to select, out of three pictures, the one which corresponded with an orally presented test item. Test items consisted of sentences or words which contained one of nine grammatical categories (NEG, SVO, GEN, PLU, POSS, PROsubj, PROobj, VERB, COPULA) being investigated in the test.

A first major research question investigated in this study was whether or not receptive grammar acquisition – like its productive counterpart – occurs in a systematic, possibly universal, fashion, and if so whether there is any similarity between the order in which grammatical items are acquired in production and the order in which they are acquired in reception.

By ordering the nine grammatical categories from the one on which the highest scores were attained to the one on which the scores were lowest, a similar rank order was found for all three L2 contact groups. However owing to issues of statistical significance, not all grammatical categories could be assigned a distinct position in the rank order. Rather four ranks could be distinguished, namely 1 NEG and SVO, 2 GEN,

3 PLU, 4 POSS, COP, PROobj, VERB, PROsubj. Within each rank the difference in mean score was not significant.

Because there is no certainty that rank orders based on mean scores in a cross-sectional study also represent acquisition orders, unless assumptions are made which may not be justified, the GT results have also been examined through implicational scaling. Both a scale with an 80% criterion of acquisition and a scale with a 90% criterion of acquisition were made. In both cases the data were found to be scalable, suggesting an implicational acquisition pattern for the nine grammatical categories in this study. The rank in the implicational scale diverged slightly from the rank order mention above, namely NEG, SVO, GEN, PLU, POSS, PROsubj, COP, VERB. However divergences only occurred in the six lowest-ranked categories, where no statistically significant differences were found. Nonetheless, the results from the implicational scales should be interpreted with care. First, as said, not all rank orders were valid because of the lack of statistical significance. Secondly, the scales were not complete because in neither case a participant was found who had acquired all nine grammatical categories. In the 90% scale the largest number of acquired categories was five out of nine, in the 80% scale seven out of nine.

Two additional factors were observed which warn for caution in drawing conclusion with regard to the rank orders and by extension with regard to the implicational scale. Although it was not possible to ascertain whether either of the above-mentioned factors significantly affected the results, and even less whether the rank order is invalid because of these factors, both factors should taken notice of.

First, it was found that semantic comprehension of the test items may have been more problematic for some categories than in other, so that some categories may have received a score which understates the participants' level of receptive grammar knowledge. Secondly, the makeup of the GT as consisting of paired test items may also have influenced the score. In some categories, half of the test items consisted of unmarked, and hence easy, test items which by themselves do not shed much light on the participants because the morpheme making the grammatical category is absent (for example, PLU: singular, in which the plural marked –s is absent), while the other half of the test items is marked, containing the grammatical item being investigated. In other categories, both test items are equally marked and equally different, with both test items

containing the grammatical item under investigation. Thus, in the former categories it may have been easier to attain a higher score than in the latter.

In addition to the route of acquisition, this study has, as a primary focus, also examined the participants' rate of receptive grammar acquisition. The participants in this study were immersion pupils who had started learning English as an L2 at the age of 5 or 6 and who were still in the early stages of the acquisition process. Both immersion education, as an increasingly popular type of bilingual education, and age of acquisition – more particularly rate of acquisition and ultimate attainment among young versus older learners – are much-debated topics in SLA.

By examining the overall GT scores, it was found that the participants' performance on the test improved as a result of L2 contact. The increase in scores in the cross-sectional study was modest but nonetheless significant when the L2 contact with 18 months of L2 contact and the group with 41 months of L2 contact were compared. A comparison of both groups with the intervening L2 contact group (29 months of l2 contact) showed that the (very modest) improvements made over the course of one year were not significant.

These results suggest a rather slow rate of acquisition among young L2 learners where the acquisition of receptive L2 grammar in an immersion setting is concerned. – although a comparison with older learners' GT results would give more insight into how slow or fast the progress in this study is. Furthermore, it would be premature to extrapolate the results from this study to L2 learners in general, in view of the focus-on-content approach and the relatively limited L2 contact in immersion settings (in comparison to many naturalistic settings) in general, and the recent implementation of the immersion programme in the Ecole Cheval Bayard in particular.

III

Conclusion

Although knowing a language entails knowing both how to produce and how to comprehend the utterances in that language, to date scholars investigating L2 grammar acquisition have focused almost exclusively on the development of productive grammar knowledge. Therefore this study has aimed at gaining insight into the dynamics of receptive second language (L2) grammar acquisition. In particular, it has been investigated whether the pattern of receptive L2 grammar acquisition, like its productive counterpart, proceeds in a relatively fixed manner.

Studies of productive L2 grammar acquisition have produced a large body of evidence that developmental patterns exist both in the consecutive acquisition of various L2 grammar features and within the acquisition of particular features. The first signs of systematicity were found in the Error Analyses conducted in the 1960s; more conclusive evidence was found in the morpheme acquisition studies conducted in the 1970s. Although morpheme studies are still considered valuable to the field of SLA, several methodological aspects are now criticized – for example the interpretation of accuracy rank orders as acquisition orders and the generalization of findings based on a limited number of morphemes to L2 grammar acquisition more generally. Furthermore morpheme studies cannot shed light on the universality of developmental patterns since all studies are limited in scope to the acquisition of English morphemes only. More recent and less disputed evidence has been found through studies of the interlanguage development of grammatical subsystems such as the verb system and syntactic structures (e.g. relativisation, question formation). Using different types of performance analysis and avoiding accuracy orders by performing implicational scaling or conducting longitudinal investigation, these interlanguage studies have successfully proven the existence of developmental orders and sequences. Consequently, although developmental patterns are not yet universally accepted, a high degree of scholarly acceptance has been reached.

On a more critical note, it may be remarked that it has often been (tacitly) assumed that the developmental patterns attested in productive L2 grammar development also pertain to receptive L2 grammar development, even though there is no empirical evidence to support this: neither the morpheme studies conducted in the 1970s nor the studies of grammatical subsystems have incorporated receptive language acquisition data in their

investigations. Rather the extrapolation of research results based on productive L2 data appears to issue from the – granted, logical – assumption that language features must be comprehended before they can be acquired, as well as on informal observations that the acquisition of receptive skills usually precedes the acquisition of productive skills – an assumption which has also led to the traditional view, in Second Language Acquisition, of receptive L2 grammar development as being necessary for and preceding the acquisition of productive L2 grammar knowledge.

Some scholars have rejected this traditional view by pointing to evidence that the acquisition of productive language skills can precede that of receptive language knowledge. Not all scholars accept this evidence, however. According to some the methodology of studies showing this primacy of language production over language reception is seriously flawed, suggesting a lack of comprehension were participants have already acquired receptive knowledge of a grammatical feature while conversely appearing to show productive grammar knowledge where participants in fact produce memorized and unanalyzed chunks. Consequently, it is far from clear whether or not the traditional views on language reception are justified, and hence it also cannot be said whether the acquisition of receptive grammar skills will indeed resemble that of productive grammar acquisition.

The aim of this study therefore was to respond to the lack of research dealing with receptive grammar acquisition. By conducting and analyzing the results a receptive grammar test it has been investigated whether receptive grammar acquisition resembles productive grammar acquisition in the existence of developmental patterns, and if so whether the developmental patterns are similar to the ones found in studies of productive grammar acquisition.

Further, to the extent that this is possible in a study of this scope, it has been examined whether the explanations that have been offered for developmental patterns in productive grammar acquisition might also apply to the results find in this study. Verifiable explanations that have been posited to account for productive L2 grammar acquisition are the frequency of occurrence of grammatical features, markedness (in the meaning this term takes in research dealing with typological language universals) or, in Pienemann's Processability Theory, the implicational acquisition order of the *processing procedures* that are required to process increasingly complex grammatical structures.

Other explanations can be found but these are either too limited in scope or not verifiable. An example of the former is the Aspect Hypothesis which accounts for developmental orders in the acquisition of verbal morphology. An example of the latter are the variegated explanations for the existence of developmental patterns which have been offered within the framework of Universal Grammar theory.

It is also worth noting that the grammar test used in this study was developed by ELIAS – a Comenius-funded European research project – to investigate the acquisition of English as an L2 in early immersion education across Europe. Likewise the data used in this study were collected, within the wider frame of this ELIAS research, in an early English immersion school in the French Community in Belgium. The participants (N=72), aged 7 to 10, were taught about 40% of their curriculum in the L2; their L2 contact at the time of testing was between 18 and 41 months. Hence the results of this study may also contribute to the debates about the age factor in SLA and about processes and outcomes in bilingual education.

Accordingly two research questions were formulated in this study. The primary one was concerned with the route of receptive grammar acquisition and its relationship to the route of productive grammar acquisition, the second one with the progress made by young L2 learners in their first years in an immersion context.

Finally, because the pattern of receptive grammar knowledge is still relatively unexplored territory, as yet no research methodologies or test materials exist which have shown their value through use in a large number of studies. With this in mind, considerable attention has been paid in this study to the advantages and disadvantages of the grammar test used.

The ELIAS grammar test used in this study contained nine grammatical categories: canonical word order (SVO), sentence negation expressed by the negator *not* (NEG), personal pronouns in the 3rd person singular, both in the nominative case (*he/she*) (PROsubj) and in the accusative case (*him/her*)(PROobj), the 3rd person singular of the possessive determiner (*his/her*) (POSS), genitive marker (–*'s*) (GEN), plural marker (–*s*) (PLU), verbal morphology of copular verbs (COP) and of full verbs (VERB) – in both cases in the 3rd person singular and plural in the simple present. In administering the grammar test, a test item containing one of these grammatical categories was orally presented to the participant, who then had to select out of three pictures the one which

corresponded with the test item. Test items were made in such a way that a correct answer was not possible unless the grammatical feature being investigated was acquired – or through successful guesswork. A semantic understanding of the test item did not suffice: for each test item two (of the three) pictures were semantically related, but only one also corresponded at the grammatical level. A third picture was a distractor which did not semantically correspond with the test item, thus revealing which test items were semantically problematic for the participants.

Ordering the mean scores which were attained in the nine distinct grammatical categories (from the one with the highest scores to one with the lowest scores) revealed the following rank order: 1 NEG and SVO, 2 GEN, 3 PLU and 4 POSS, COP, PROobj, VERB and PROsubj. To avoid making the same mistake as the morpheme studies (see above) this rank order was not interpreted as an acquisition order. Rather implicational scaling was used to determine whether an implicational acquisition order was likely to exist. The nine grammatical categories were found to be scalable both in case of an 80% criterion for acquisition and in case of a 90% criterion for acquisition. From this it could be concluded that the nine grammatical categories are acquired in an implicational order.

Importantly, a number of factors were identified which warned against interpreting these results as conclusive evidence for the existence of developmental patterns in receptive L2 grammar acquisition. A first factor was the incompleteness of the implicational scales: none of the participants in this study had acquired all nine grammatical categories. As a second potentially problematic factor, it was observed that certain test items were more difficult to semantically process than others. Although an examination of this factor showed no significant effect on the rank order in this study, it is nonetheless a factor which should not be discarded too readily. Thirdly, the grammatical items in the ELIAS GT were paired, with every pair of test items containing two alternatives of the grammatical item under investigation (e.g. PROobj: *masculine* and PROobj: *feminine*). An analysis of these alternatives showed that in some categories one of the alternatives was distinctly less marked and hence much easier than the other (e.g. PLU: *singular* versus PLU: *plural*). In other categories (such as PROobj above), both alternatives of the grammatical category were equally marked. Although, like semantic processing, this factor did not appear to have had a significant effect of the rank order, this factor too cannot be altogether discarded.

Likewise, the results from this study should not be extrapolated too readily to receptive L2 grammar acquisition in general. Results may be different for other grammatical categories, for participants learning the L2 in other contexts and for older L2 learners. Furthermore, although no conspicuous L1 effects were identified in this study, it is possible that L1 background affects the route of receptive grammar development, either through positive or negative transfer.

As to the relationship between receptive and productive grammar acquisition, the rank order in this study was found to be divergent from the rank orders identified in earlier studies of productive L2 grammar acquisition. Consequently it is also highly unlikely that explanations such as frequency of occurrence and markedness pertain to the rank order found in this study. Processability Theory has not been tested because not all processing procedures were represented in this study.

With regard to the progress made by young immersion pupils in their first years of L2 contact, a cross-sectional analysis of the results attained in three L2 contact groups (18, 29 or 41 months of L2 contact) appeared to suggest a relatively slow rate of acquisition, thereby supporting scholarly opinions which argue that, while young L2 learners may be more likely to attain native-like levels of proficiency, they do not necessarily 'absorb the L2 grammar like sponges'. However, the results may also be owing to the learning setting. Studies of immersion education have shown that the acquisition of grammar acquisition is impeded by the focus-on-meaning approach taken in immersion education. Likewise the cross-sectional design of the study and the fact that the immersion programme has only fairly recently been implement in the participants' school should be considered as factors influencing the test results. Therefore a comparison with older L2 learners and with L2 learners in other learning settings is required for more conclusive evidence.

To conclude, the challenge for future researchers interested in receptive grammar acquisition lies in developing tests and methodologies which control for semantic and grammatical processing, which investigate a wider range of grammatical items – on the level of morphology and syntax – and which allow to test theories and hypotheses such as Processability Theory and frequency of occurrence. In addition, research will have to be conducted in a wide variety of settings so that the influence of factors such as age, L2 learning setting and L1 background can be assessed.

In essence, the same challenges that have had to be faced, and still have to be faced, by researchers dealing with productive L2 grammar acquisition will have to be faced by researchers taking on the task of investigating receptive L2 grammar acquisition.

References

Aitchison, J. (1998). *The Articulate Mammal. An Introduction to Psycholinguistics* (4th ed.). London/New York: Routledge.

Andersen, R. & Shirai, Y. (1996). Primacy of Aspect in First and Second Language Acquisition: The Pidgin/Creole Connection. In W. C. Ritchie & T. K. Bhatia (eds.), *Handbook on Language Acquisition* (pp. 527-570). San Diego, CA: Academic Press.

Bailey, N., Madden, C., & Krashen, S. (1974). Is There a Natural Sequence in Adult Second Language Learning? *Language Learning 24,* 235-243.

Baker, C. (2006). *Foundations of Bilingual Education and Bilingualism* (4th ed.). Clevedon, England: Multilingual Matters.

Baker, C., & Prys Jones, S. (1998). *Encyclopedia of Bilingualism and Bilingual Education*. Clevedon, England: Multilingual Matters.

Bardova-Harlig, K. (1999). From Morpheme Studies to Temporal Semantics. *Studies in Second Language Acquisition, 21,* 341-382.

Bates, E., Dale, P. S., Thal, D. (1995). Individual Differences and their Implications for Theories of Language Development. P. Fletcher & B. MacWhinney (Eds.), *Handbook of Child Language* (pp. 96-151). Oxford: Blackwell Publishing.

Berthele, R., Kaiser, I., & Peyer, E. (in press.). Receptive Learner Varieties. *Language Learning*.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Ltd.

Brown, R. (1973). *A First Language*. London: George Allen & Unwin Ltd.

Clahsen, H. (1992). Learnability Theory and the Problem of Development in Language Acquisition. In J. Weissenborn, H. Goodluck & T. Roeper (Eds.), *Theoretical Issues in Language Acquisition. Continuity and Change in Development* (pp. 53-76). NJ: Lawrence Erlbaum Associates.

Cook, V. (1993). *Linguistics and Second Language Acquisition*. Basingstoke, U.K.: The Macmillan Press Ltd.

Corder, S. P. (1967). The Significance of Learners' Errors. *International Review of Applied Linguistics, 5,* 161-159.

Corder, S. P. (1981). *Error Analysis and Interlanguage*. Oxford: Oxford University Press.

Crystal, D. (2003). *A Dictionary of Linguistics and Phonetics* (5th ed.). Malden, MA: Blackwell Publishing.

Cummins, J. (2000). *Immersion Education for the Millennium: What We Have Learned from 30 Years of Research on Second Language Immersion*. Retrieved October 8, 2008 from http://www.iteachilearn.com/ cummins/immersion2000.html

de Graaff, R., & Housen, A. (2009). Investigation the Effects and Effectiveness of L2 Instruction. In M. H. Long & C. J. Doughty (Eds.), *The Handbook of Language Teaching*, pp. 726-755. Malden, MA: Wiley-Blackwell.

de Hoop, H., & Krämer, I. (2006). Children's Optimal Interpretations of Indefinite Subjects and Objects. *Language Acquisition, 13*(2), 103-123.

de Villiers, J. & de Villiers, P. (1973). A Cross-sectional Study of the Acquisition of Grammatical Morphemes in Child Speech. *Journal of Psycholinguistic Research 2,* 267-278.

Di Biase, B., & Kawaguchi, S. (2002). Exploring the Typological Plausibility of Processability Theory: Language Development in Italian Second Language and Japanese Second Language. *Second Language Research, 18*(3), 274-302.

Dulay, H., & Burt, M. (1974). Natural Sequences in Child Second Language Acquisition. *Language Learning, 24,* 37-53.

ELIAS (2008). *Early Language and Intercultural Acquisition Studies*. Retrieved February 25, 2009 from www.elias.bilikita.org

Ellis, N. (1998). Emergentism, Connectionism, and Language Learning. *Language Learning, 48,* 631-664.

Ellis, N. (2002). Frequency Effects in Language Processing: A Review with Implications for Theories of Implicit and Explicit Language Acquisition. *Studies in Second Language Acquisition, 25,* 143-188.

Ellis, N. (2005). Constructions, Chunking, and Connectionism: The Emergence of Second Language Structure. In C. J. Doughty, & M. H. Long (Eds.), *The Handbook of Second Language Acquisition,* pp. 63-103. Malden, MA: Blackwell Publishing.

Ellis, R. (2008). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

Felix, S. (1992). Language Acquisition as a Maturational Process. In J. Weissenborn, H. Goodluck & T. Roeper (Eds.), *Theoretical Issues in Language Acquisition. Continuity and Change in Development* (pp. 25-52). NJ: Lawrence Erlbaum Associates.

Gass, S., & Selinker, L. (2008). *Second Language Acquisition: An Introductory Course* (3rd ed.). New York: Routledge.

Genesee, F. (2004). What Do We Know About Bilingual Education for Majority Language Students? In T. K. Bhatia & W. Ritchie (Eds.), *Handbook of Bilingualism and Multiculturalism* (pp. 547-576). Malden, MA: Blackwell.

Goldschneider, J. M. & De Keyser, R. M. (2001). Explaining the "Natural Order of L2 Morpheme Acquisition" in English: A Meta-analysis of Multiple Determinants. *Language Learning 55*(1), 27-77.

Grevisse, M., & Goosse, A. (2007). *Le bon usage*. (14th ed.). Paris: Groupe De Boeck.

Hakuta, K. (1976). A Case Study of a Japanese Child Learning English as a Second Language. *Language Learning, 26,* 321-51.

Hatch, E. (1983). *Psycholinguistics: A Second Language Perspective*. Rowly, Massachusetts: Newbury House Publishers, Inc.

Hatch, E., & Lazaraton, A. (1991). *The Research Manual. Design and Statistics for Applied Linguistics*. Boston, MA: Heinle & Heinle Publishers.

Hatch, E., & Farhady, H. (1982). *Research Design and Statistics for Applied Linguistics*. Rowly, MA: Newbury House Publishers, Inc.

Hendriks, P. & Koster, C. (in press). Production/Comprehension Asymmetries in Language Acquisition. *Lingua,* doi:10.1016/j.lingua.2010.02.002

Hendriks, P. (2007). A Unified Explanation for Production/Comprehension Asymmetries in Language Acquisition. In A. Gavarró and M. João Freitas (Eds.). *Generative Approaches to Language Acquisition* (pp. 240-251). Cambridge Scholars Publishing.

Hendriks, P., de Hoop, H., & Lamers, M. (2005). Asymmetries in Language Use Reveal Asymmetries in the Grammar. P. Dekker & M. Franke (Eds.), *Proceedings of the Fifteenth Amsterdam Colloquium* (pp. 113-118). Amsterdam: ILLC.

Hurewitz, F., Brown-Schmidt, S., Thorpe, K., Gleitman, L. R., & Trueswell, J. C. (2000). One Frog, Two Frog, Red Frog, Blue Frog: Factors Affecting Children's Syntactic Choices in Production and Comprehension. *Journal of Psycholinguistic Research, 29*(6), 597-626.

Hyltenstam, K. (1984). The Use of Typological Markedness Conditions as Predictors in Second Language Acquisition: The Case of Pronominal Copies in Relative Clauses. In A. Andersen (Ed.), *Second Languages: A Cross-Linguistic Perspective* (pp. 39-58). Rowly, MA: Newbury House Publishers.

Hyltenstam, K. (1986). Markedness, Language Universals, Language Typology, and Second Language Acquisition. In C. F. Pfaff (ed.), *First and Second Language Acquisition Processes* (pp. 53-78). Cambridge: Newbury House Publishers.

James, C. (1998). *Errors in Language Learning and Use*. London/New York: Longman Ltd.

Kaplan, R. M., & Bresnan, J. (1982). Lexical-Functional Grammar: A Formal System for Grammatical Representation. In J. Bresnan (ed.), *The Mental Representation of Grammatical Relations* (pp. 173-282). Cambridge (Massachusetts): The Massachusetts Institute of Technology.

Keenan, J. M., & MacWhinney, B. (1987). Understanding the Relationship between Comprehension and Production. In H. W. Dechert & M Raupach (Eds.), *Psycholinguistic Models of Production* (pp. 149-155). Norwood, N.J.: Ablex Publishing Corporation.

Kellerman, R. (1984). The Empirical Evidence for the Influence of the L1 in Interlanguage. In A. Davies, C. Criper & A. P. R. Howatt (eds.), *Interlanguage* (pp. 98-122). Edinburgh: University Press.

Kersten, K., Frey, E., & Hähnert, A. (Eds.). (2009). Early Language and Intercultural Acquisition Studies: Progress Report. Magdeburg: ELIAS

Klein, W. (1995). The Acquisition of English. In R. Dietrich, W. Klein & C. Noyau (eds.), *The Acquisition of Temporality in a Second Language* (pp. 31-70). Amsterdam/Philadelphia: John Benjamins Publishing Co.

Krashen, S. (1981). *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon Press.

Krashen, S. (1985). *The Input Hypothesis: Issues and Implications*. London/New York: Longman.

Larsen-Freeman, D. (1975). The Acquisition of Grammatical Morphemes by Adult ESL Learners. *TESOL Quarterly 9,* 409-430.

Larsen-Freeman, D. (1976). An Explanation for the Morpheme Acquisition Order of Second Language Learners. *Language Learning 26*, 125-134.

Larsen-Freeman, D. (2002). Making Sense of Frequency. *Studies in Second Language Acquisition 24*(2), 275-282.

Lightbown, P. (1984). The Relationship between Theory and Method in Second Language Acquisition Research. In A. Davies, C. Criper, and A. Howatt (eds.), *Interlanguage* (pp. 241-252); Edinburgh: Edinburgh University Press.

Lightbown, P. (1985). Can Language Acquisition be Altered by Instruction? In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and Assessing Second Language Acquisition,* pp. 101-112. Clevedon, England: Multilingual Matters.

Lightbown, P., & Spada, N. (1993). *How Languages Are Learned*. Oxford: Oxford University Press.

Long, M. H., & Sato, C. J. (1984). Methodological Issues in Interlanguage Studies. In A. Davies, C. Criper & A.P.R. Howatt (eds.), *Interlanguage* (pp. 253-279). Edinburgh University Press.

Ortega, L. (2009a). Sequences and Processes in Language Learning. In M. H. Long & C. J. Doughty (Eds.), *Handbook of language teaching* (pp. 81-105). Malden, MA: Wiley-Blackwell.

Ortega, L. (2009b). *Understanding Second Language Acquisition*. London: Hodder Education.

Pienemann, M. (1985). Learnability and Syllabus Construction. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and Assessing Second Language Acquisition,* pp. 23-76. Clevedon, England: Multilingual Matters.

Pienemann, M. (1998). *Language Processing and Second Language Development. Processability Theory*. Amsterdam: John Benjamins Publishing Co.

Pienemann, M. (2005a). An introduction to Processability Theory. In M. Pienemann (ed.), *Cross-Linguistic Aspects of Processability Theory* (pp. 1-60). Amsterdam: John Benjamins Publishing Co.

Pienemann, M. (2005b). Language Processing Capacity. In C. J. Doughty & M. H. Long (eds.), *The Handbook of Second Language Acquisition* (pp. 679-714). Malden, MA: Blackwell Publishing.

Pienemann, M., & Håkansson, G. (1999). A Unified Approach Towards the Development of Swedish as L2: A Processability Account. *Studies in Second Language Acquisition,* 21, 383-420.

Pica, T. (1983). Adult Acquisition of English as a Second Language Under Different Conditions of Exposure. *Language Learning, 33,* 465-497.

Rickford, J. R. (2002). Implicational Scales. In J. K. Chambers, P. Trudgill, & Natalie Schilling-Estes (Eds.), *The Handbook of Language Variation and Change,* pp. 142-167. Malden, MA: Blackwell Publishing Ltd.

Robison, R. E. (1995). The Aspect Hypothesis Revisited: a Cross-Sectional Study of Tense and Aspect Marking in Interlanguage. *Applied Linguistics 16*, 344-370.

Rohde, A. (1996). The Aspect Hypothesis and the Emergence of Tense Distinctions in Naturalistic L2 Acquisition. *Linguistics, 34*(5), 1115-1137.

Rosansky, E. (1976). Methods and Morphemes in Second Language Acquisition. *Language Learning, 26,* 409-425.

Sakai, H. (2008). An Analysis of Japanese University Students' Oral Performance in English Using Processability Theory. *System, 36,* 534-549.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics, 10,* 219-231.

Selinker, L. (1992). *Rediscovering Interlanguage*. Harlow, Essex: Longman Group UK Limited.

Slobin, D. I. (1973). Cognitive Prerequisites for the Development of Grammar. In C. A. Ferguson and D. I. Slobin (eds.), *Studies of Child Language Development* (pp. 175-208). New York: Holt, Rinehart & Winston Inc.

Spenader, J., Smits, E.-J., & Hendriks, P. (2009). Coherent discourse solves the Pronoun Interpretation Problem. Poster presented at the KNAW Academy Colloquium on Language Acquisition and Optimality Theory. Amsterdam.

Swain, M. (1995). Three Functions of Output in Second Language Learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and Practice in Applied Linguistics* (pp. 125-143). Oxford: Oxford University Press.

Tasseva-Kurktchieva, M. (2008). What About Grammar? Comprehension and Production at the Initial Stage of L2 Acquisition. R. Slabakova et al. (Eds.), *Proceedings of the 9th Generative Approaches to Second Language Acquisition Conference* (pp. 242-250). Somerville, MA: Cascadilla Proceedings Project.

Unsworth, S. (2007). L1 and L2 Acquisition between Sentence and Discourse: Comparing Production and Comprehension in child Dutch. *Lingua, 117,* 1930-1958.

VanPatten, B. (2002). Processing Instruction: An Update. *Language Learning 52*(4), 755-803.

White, L. (2005). On the Nature of Interlanguage Representation: Universal Grammar in the Second Language. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 19-42). Malden, MA: Blackwell Publishing.

Appendices

Appendix A Grammar Test Score Sheets (Including Test Items)

ELIAS GRAMMAR TEST A: SCORES

Name (Last): **(First):**
❑ Female ❑ Male **ID-No.:** **Preschool:**

Date of Testing: Year Month Day
Date of Birth:

Age in years and completed months: Year Month Day
Duration of L2-contact:

	Prompt	Correct	Error	Distractor	Comments
Training	dog	(A)	(B,C)	-	
Training	cat	(C)	(B,A)	-	
1	the boy is touching the girl	(A)	(C)	(B)	
2	cats	(A)	(C)	(B)	
3	the girl is kissing her	(C)	(A)	(B)	
4	his cat	(C)	(B)	(A)	
5	the boy is not running	(C)	(B)	(A)	
6	the sheep eats	(A)	(C)	(B)	
7	the deer are white	(A)	(B)	(C)	
8	the girl is kissing the boy	(C)	(A)	(B)	
9	he is singing	(C)	(A)	(B)	
10	the dog is chasing the cat	(B)	(C)	(A)	
11	pen	(C)	(B)	(A)	
12	the father is chasing him	(C)	(A)	(B)	
13	her apple	(A)	(B)	(C)	
14	the dog is sleeping	(A)	(B)	(C)	
15	the fish jump	(B)	(C)	(A)	
16	the fish is black	(A)	(B)	(C)	
17	the girl is feeding the boy's dog	(C)	(A)	(B)	
18	she is drinking	(B)	(C)	(A)	
19	the fish is kissing the frog	(A)	(B)	(C)	
20	balls	(B)	(C)	(A)	
21	the mother is touching her	(B)	(A)	(C)	
22	his ball	(A)	(C)	(B)	
23	the duck is not eating	(C)	(B)	(A)	
24	the reindeer sleeps	(C)	(A)	(B)	
25	the sheep are happy	(B)	(A)	(C)	
26	the girl is carrying the boy	(B)	(C)	(A)	
27	he is running	(B)	(C)	(A)	

ELIAS GRAMMAR TEST B: SCORES

Name (Last): **(First):**
❑ Female ❑ Male **ID-No.:** **Preschool:**

	Year	Month	Day		Year	Month	Day
Date of Testing:				Age in years and completed months:			
Date of Birth:				Duration of L2-contact:			

	Prompt	Correct	Error	Distractor	Comments
Training	duck	(B)	(A,C)	-	
Training	fish	(A)	(B,C)	-	
1	the girl is touching the boy	(C)	(A)	(B)	
2	cat	(B)	(C)	(A)	
3	the girl is kissing him	(A)	(C)	(B)	
4	her cat	(B)	(C)	(A)	
5	the boy is running	(B)	(C)	(A)	
6	the sheep eat	(C)	(B)	(A)	
7	the deer is white	(B)	(C)	(A)	
8	the girl is kissing the boy's dog	(A)	(C)	(B)	
9	she is singing	(A)	(C)	(B)	
10	the cat is chasing the dog	(C)	(B)	(A)	
11	pens	(A)	(C)	(B)	
12	the father is chasing her	(A)	(C)	(B)	
13	his apple	(B)	(A)	(C)	
14	the dog is not sleeping	(A)	(B)	(C)	
15	the fish jumps	(A)	(C)	(B)	
16	the fish are black	(B)	(C)	(A)	
17	the girl is feeding the boy	(A)	(C)	(B)	
18	he is drinking	(C)	(B)	(A)	
19	the frog is kissing the fish	(B)	(A)	(C)	
20	ball	(C)	(A)	(B)	
21	the mother is touching him	(A)	(B)	(C)	
22	her ball	(C)	(A)	(B)	
23	the duck is eating	(A)	(B)	(C)	
24	the reindeer sleep	(B)	(A)	(C)	
25	the sheep is happy	(A)	(C)	(B)	
26	the girl is carrying the boy's dog	(C)	(B)	(A)	
27	she is running	(C)	(B)	(A)	

Appendix B Example: Multiple Choice Options for PLU (Test Item *Cats*)