INVESTIGATING THE ACQUISITION OF THE GERMAN PASSIVE BY ADULT LEARNERS WITH L1 ENGLISH IN A FORMAL LEARNING ENVIRONMENT:

A LEARNER CORPUS APPROACH

Sheena Caroline Elizabeth Thoday

Submitted for the degree of Doctor of Philosophy

Heriot-Watt University

School of Management and Languages

November 2014

ABSTRACT

This cross-disciplinary study draws on theories of linguistics, second language acquisition and language pedagogy in order to investigate the acquisition of the German passive by learners in a formal learning environment whose first language is English. This construction has hitherto received little attention from researchers, possibly because of the complexity of the available means of expressing it. The study focuses on the two periphrastic syntactic alternatives, each of which has a different semantic interpretation.

With the general aim of identifying patterns in the learners' usage of the construction over time, both synchronically and diachronically-collected data are assembled in a learner corpus and examined in a multi-method, multi-level study which differs from the single case study approach to language acquisition investigations which has often prevailed in the past. A quantitative, electronically-driven analysis is followed by a qualitative, judgement-based commentary. A comparative analysis of data from native speakers of German adds a further dimension to the discussion.

Results of the analyses indicate differences between the written production of the two research populations and variation in the sequence of development and usage of the two German passives in the formal learning context which, it is suggested, may in part be due to the role played by instruction.

DEDICATION

To my family and to lifelong learning

ACKNOWLEDGEMENTS

rewarding. In comparison, my association with Dr Maggie Sargeant has been relatively short but no less valuable. I have very much appreciated her incisive and supremely fair commentaries and her attention to detail, not to mention her boundless optimism. I feel enormously privileged to have been associated with them both. Mike and Maggie have had the unenviable task of combining the roles of colleague and friend with that of supervisor and I owe them a debt of gratitude for always striking the best balance.

Of course, without support from my family I would never have started this study. Sarah and Hannah and their husbands have never stopped encouraging me in these endeavours. Keith's constant and invaluable contribution, including the countless cups of tea, spurred me on; he has gallantly done his own thing in recent months so that I could do mine. Without his help and advice along the way I would never have reached the end. I am indebted to you all for believing in me. Thank you.

ACADEMIC REGISTRY
Research Thesis Submission

Name*:*	SHEENA CAROLINE ELIZABETH THODAY		
School/PGI:	SCHOOL OF MANAGEMENT AND LANGUAGES		
Version: *(i.e. First, Resubmission, Final)*	FINAL	Degree Sought (Award **and** Subject area)	DOCTOR OF PHILOSOPHY SECOND LANGUAGE ACQUISITION

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

Signature of Candidate*:*		Date:	

Submission

Submitted By *(name in capitals):*	
Signature of Individual Submitting:	
Date Submitted:	

For Completion in the Student Service Centre (SSC)

Received in the SSC by *(name in capitals):*			
Method of Submission *(Handed in to SSC; posted through internal/external mail):*			
*E-thesis Submitted (**mandatory for final theses**)*			
Signature:		Date:	

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GLOSSARY OF TERMS

Communicative Language Teaching (CLT): Teaching a language with the primary focus on the transmission of meaning but with some attention to grammatical form either through Focus on Form or constructive feedback (Spada 2007).

Contrastive Interlanguage Analysis (CIA): Comparing and contrasting learner language not only in two different languages but also in learner and native varieties of the same language (Granger 1996: 43).

Cross-linguistic Influence (CLI): Carrying over patterns both from a learner's L1 to L2 and from L2 to L1 in cases of first language attrition (Sharwood Smith 1998) (see also **Positive Transfer** and **Negative Transfer**).

Emergence Criterion: Defining acquisition as the moment at which a new construction or stage of development has emerged independently in a learner's production on a few occasions as opposed to when the relevant constructions have been fully mastered. '[T]he point in time at which certain skills have, in principle, been attained or at which certain operations can, in principle, be carried out. (…) [T]his is the beginning of an acquisition process' (Pienemann 1998: 138).

Focus on Form: 'an occasional shift of attention to linguistic code features – by the teacher and/or one or more students - triggered by perceived problems with comprehension or production' (Long 1998: 23).

Formulaic Language: 'a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar' (Wray 2002: 9).

Negative Transfer: Carrying over patterns from a learner's L1 which results in incorrect usage in the L2. Negative transfer inhibits acquisition (Isurin 2005).

Overuse: use of a linguistic item by a NNS more frequently than it would be used by a NS in a similar context (Gilquin & Paquot 2008).

Positive Transfer: Carrying over patterns from a learner's L1 which results in correct usage in the L2. Positive transfer facilitates acquisition (Isurin 2005).

Underuse: use of a linguistic item by a NNS less frequently than it would be used by a NS in a similar context (Gilquin & Paquot 2008).

ABBREVIATIONS

ACC	Accusative case
ANNIS	Annotation of Information Structure
AUX	Auxiliary verb
CA	Contrastive Analysis
CAH	Contrastive Analysis Hypothesis
CEFR	Common European Framework of Reference for Languages
CHAT	Codes for the Human Analysis of Transcripts
CHILDES	Child Language Data Exchange System
CIA	Contrastive Interlanguage Analysis
CLAN	Computerized Language Analysis
CLI	Cross-linguistic Influence
CLT	Communicative Language Teaching
DaF	Deutsch als Fremdsprache
DAT	Dative case
DET	Determiner
ESL	English as a Second Language
FALKO	Fehlerannotiertes Lernerkorpus
FDH	Fundamental Difference Hypothesis
FFI	Form-focused Instruction
FL	Foreign Language
FLLOC	French Learner Language Oral Corpora
GEN	Genitive case
GJE	Grammaticality Judgement Exercise
ICLE	International Corpus of Learner English
IH	Interface Hypothesis
IL	Interlanguage
INFIN	Infinitive
L1	First language
L1 E	English as mother tongue/first language
L1 G	German as mother tongue/first language
L2	Second language
LFG	Lexical Functional Grammar
LINCS	Languages & Intercultural Studies

MOGUL	Modular On-line Growth and Use of Language
NOA	Needs Only Analysis
NOM	Nominative case
NS	Native Speaker
NNS	Non-native Speaker
PI	Processing Instruction
PoS	Part of Speech
PP	Past Participle
PT	Processability Theory
SLA	Second Language Acquisition
SPLLOC	Spanish Learner Language Oral Corpora
TBL	Task-based Learning
TL	Target Language
TTR	Type/Token Relationship
T-unit	The shortest sentence or independent clause which can stand on its own
UCL	Université Catholique de Louvain
UG	Universal Grammar
UK	United Kingdom
USA	United States of America

1pers	first person
2pers	second person
3pers sing	third person singular
3pers pl	third person plural
pres	present tense
imp	imperfect tense

PUBLICATION

Thoday, E. (2008) 'Issues in building learner corpora: an investigation into the acquisition of German passive constructions' in *Newcastle Working Papers in Linguistics*, 14, 145-155.

CHAPTER 1 - INTRODUCTION

1.1   The Background to the Study

The cross-disciplinary study described in this thesis investigates the acquisition of a complex construction, the German passive, by learners whose first language is English. The environment in which these learners are studying is formal and academic: they are students of languages at a British university. Both synchronically and diachronically collected data, assembled in a learner corpus as defined in Chapter 2.6.3.1, are examined in a multi-method, multi-level study which differs from the single case-study approach which has often been used in investigating the acquisition of German passives in the past. A learner corpus provides a rich supply of data on which to base comments on L2 acquisition; it does not itself contribute to the development of theoretical considerations. This study aims to offer observations on the acquisition of the German passive by second language learners whose first language is English in a specific learning environment, with reference to a number of theories of grammatical development, rather than to test a specific theory.

The early observations which provided the foundation for this study arose from pedagogical experience of teaching the German passive to learners with English as a first language (L1 E).[1] It became apparent that despite competent, accurate production of the passive in form-targeted class exercises during which the learners' attention was specifically drawn to a particular paradigm, the same accuracy was not displayed in free-writing tasks with no time constraints when learners would presumably have recourse to whatever formal knowledge about the grammar of the passive construction was at their disposal. Nor was accuracy demonstrated in all cases in examination tests carried out with limited time, when each of several grammar exercises was focused on a different form. In addition, variation in the production of individual learners over time was noted; some made progress, others did not and in some cases *re*gression was evident. Of the two syntactic forms of the German passive, the *sein* passive appeared to be produced not only less frequently but less accurately.

---

[1] This term will be explained in Section 1.2.

These interesting observations led to the decision to conduct an in-depth study into the acquisition of the German passive by native speakers of English in the context of a formal learning environment, in other words as opposed to studying a second language acquired during everyday contact with users of a foreign language. For the purposes of this study, it was decided to focus on the written language of learners in a formal learning context who are classified as adults both by their age and by the level of their academic qualifications which categorise them as relatively advanced learners.[2] The written language which provides the data for the study is collected, stored and, partially at least, analysed electronically. It is only comparatively recently that such electronic collections have been made possible, by the technological advances following the development of the World Wide Web in the late 1980s.

In the time that has elapsed since the beginning of that decade, teaching and learning methods have also changed irrevocably. Whereas language teachers used to rely on methods in textbook format and learners submitted all text work in handwritten form, it is now possible for teachers to access pedagogical materials from the internet and for learners to submit written exercises electronically and receive feedback the same way, whether it is from a tutor in person or automated by a computer program. Not only teaching and learning have been affected by these advances; it is within the same time-frame that the type of research study which is described here has emerged: Second Language Acquisition (SLA) research has its foundations in the 1970s and research based on electronic collections of learner language, learner corpora, was first described in an edited volume of articles in the late 1990s (Granger 1998). The research aims of the study reported here, which will be given in Section 1.7, are intended to lead to a contribution to both these relatively young areas of research.

Having given an outline of the pedagogical and typological background to this study, the subject-specific terms which will recur throughout it will be explained in the next section of this introductory chapter. The study will then be set within its cross-disciplinary theoretical and methodological contexts and subsequently the research aims and the research questions which the study seeks to answer will be provided. Naming and formatting conventions to be used in the thesis are then explained and the structure of the whole is outlined in the concluding section of the chapter.

---

[2] This term will be explained in Section 1.2.

1.2 Introducing the Concepts

The terms *corpus* and *cohort* are used in this study in the following way: the latter refers to a collection of individuals and the former to a collection of texts. The Languages and Intercultural Studies (LINCS) corpus which provides the data for this study is composed of three subsections, A, B, and C, all of which contain written texts in German. With the exception of one respondent, A7, the texts in the A corpus were produced by a student cohort, the A cohort, whose mother tongue is English (English is the first language they learnt in childhood; they are native speakers (NSs) of English). A7 is a native speaker (NS) of German who studied with the A cohort. She therefore falls into two categories, as explained further in Section 1.4. The B corpus comprises texts written by students whose mother tongue is German (NSs of German). The third element, the C corpus, is a small 'input corpus' containing a collection of formal texts of different types which were used for pedagogical purposes in the students' training.

Cook's differentiation between *learner* and *user* should be mentioned in the pedagogical context of the present study. In his terms, the former denotes any individual who is learning a second language[3] and the latter, one who uses that language outside a classroom situation (Cook 2008: 12). He acknowledges that the terms can be synonymous; an individual can be both a *learner* and a *user* if, for example, she is learning a second language (L2) for the purposes of using it on a daily basis in the country where she is currently residing. This was the case for all of the respondents in the A cohort during their time abroad, when they were both learning and using their L2. For the purposes of this study, only respondent A7 falls into this category, since she is learning and using English as her second language, while simultaneously being an NS user of German, as explained in the next paragraph. It could also be said that she is a learner of her mother tongue, too, to the extent that an individual's language is in a constant state of development, depending on circumstances and language change itself. It should be noted that Cook points out (Cook 2008: 12) that describing an individual who has been residing in a foreign country for years and using its language competently on a daily basis as a learner is less than complimentary. The term *learner* carries no negative connotations in this study; it is used here simply to differentiate the A cohort

---

[3] Cook's definition in fact refers to 'acquiring a second language' (Cook 2008: 11). The terms 'acquisition' and 'learning' will be discussed in Chapter 2.

from the B cohort. The term *user* is applied to an NS of the language which was learnt in childhood (identified as 'mother tongue' above); this language is the user's **L1.**

It follows that the *learners* here are the respondents in the A cohort of students. The majority of these students have English as L1 and are *learners* of German as L2. The respondents in the B corpus are all L1 German *users* (L1 G). One respondent (A7) occupies a particular place in the study as a representative in both the B corpus and the A corpus; she is an L1 *user* of German and an L2 *learner* of English. As far as her education at tertiary level is concerned she is 'learning' (studying) with the other respondents in the A cohort.

In the present context, *foreign language* (FL) refers to a language other than the learner's L1 and is synonymous with *second language* (L2), which might be the second or third or more in the learner's repertoire.[4] *Second Language Acquisition* (SLA) refers to the research discipline itself. For the purposes of this study, German is named as the learners' L2, where L2 denotes *a* foreign language being learned. In fact, all the students in the A cohort studied two foreign languages in their curriculum; they were expected ultimately to reach the same standard in French or Spanish as in German, but it was not established from their educational records which FL was the first on which they embarked. In this respect we do not follow Cook (2008: 11), who defines foreign language as 'for long-term future use in other countries' and second language as 'for immediate use within the same country'. Both expressions refer in this study to the language(s) being studied in the academic context.

The items which constitute the LINCS corpus described here are the products of a *formal learning environment*, a classroom environment where learning is promoted through contact with the FL in spoken and written form and supported by grammar instruction. In Cook's terms, this equates to the 'academic style' of teaching (Cook 2008: 237), an expression which is particularly apt here since the study focuses on learner language produced in an academic context. At the institution where this research was carried out, applied language studies combine the teaching of language according to the communicative method with more formal grammar instruction,

---

[4] Cook (2008: 11) has proposed the term 'multi-competence' to describe the ability of such users; other researchers prefer the term 'bilingual' when referring to individuals who have capabilities in more than one language, using the term to include multilingualism and varying levels of language ability (e.g. Romaine 1996).

intended for those who have an intellectual interest in the forms of the language they are studying as well as use of the language itself, rather than for those who require useful informal phrases for tourist purposes, for example. The type of environment in which this teaching occurs places importance on the role of pedagogical support and thus differs from the one in which naturalistic acquisition takes place, which is enabled through engagement with native speakers in everyday situations, without explicit instruction. Learners of a second language in a classroom situation have the expertise of teachers, both as trainers and as users of the FL themselves, to guide them towards acquisition. In contrast, learners in a naturalistic setting acquire their second language via 'interaction with peers' (Lightbown & Spada 2011: 109). These learners may be adults at work or children at school who are hearing the second language in an instructional setting rather than being taught how to use it.

1.3   The Linguistic Context

An enquiry into linguistic variation in German in contrast to English is interesting because the two languages share a common linguistic ancestry as part of the West-Germanic subsection of the Indo-European language group. Hawkins (1986) examines some of the differences between them, brought about by diachronic language change. English inflectional morphology, for instance, has shrunk over time in comparison to German (few examples of the case system remain in English) and Hawkins suggests that this tendency to simplification extends also to syntax, which is the focus here. German offers, for example, two levels of specificity in order to differentiate between statal passives and actional passives[5] – each requires the use of a different auxiliary, *sein* (= be) or *werden* (= become) respectively. A brief description of the syntactic and semantic differences between the German passive and the English passive is included at this juncture for the sake of clarity, but a more comprehensive comparison, carried out with specific reference to the pedagogical context of this study, is the focus of Chapter 3.

The *werden* passive describes the action in the process of being carried out (1a) whereas the *sein* passive expresses a state following from an action which has already occurred

---

[5] These terms are defined in Chapter 2.2.2.2.

(2a). In contrast, just one auxiliary verb in English (be) is used to express both passives (1c and 2c).[6]

(1a)	Der Tisch	wird	gedeckt.
(1b)	The table	becomes/is becoming (pres)	set.
(1c)	The table is (in the process of) being set.		

(2a)	Der Tisch	ist	gedeckt.
(2b)	The table	is (pres)	set.
(2c)	The table is (i.e. has already been) set.[7]		

(Durrell 1996: 299)

Hawkins makes an observation which is particularly useful for this study: 'Roughly, where the morphological and syntactic rules of English and German contrast, there is less correspondence between surface form and semantic representation in English' (Hawkins 1986: xiii). Based on the example of the German syntactic passive in the previous paragraph, it should be expected that grammatical form and semantic meaning of the passive would differ in the two languages. This is indeed the case, as will be shown in the discussion which follows in Chapter 3. German does give 'a "tighter fit" between surface form and semantic representation' (Hawkins 1986: 122), and it is the tension between the two which may account for acquisitional idiosyncracies in learners with L1 E.

1.4   The Pedagogical Context

The contributors to the A section of the LINCS corpus are termed *adult* learners in this study. They have already obtained university entrance qualifications and were between the ages of 18 and 20 on admission to university. More detailed profiles of these learners are presented in Chapter 4.6.1.

---

[6] The verb 'get' (*bekommen* in German) occurs less frequently. The use of this form is largely colloquial and is discussed in Chapter 3, together with other syntactic means of expressing the passive in German.
[7] It is often the case that a German present tense *sein* passive is understood as an English past tense (2c). This can be explained by the fact that the German sentence is close in meaning to a perfect tense e.g. (2a) is related to *Der Tisch ist gedeckt worden* (perf) (The table has been set) (Durrell 1996: 299).

The course of FL study which the A cohort followed is of 4 years' duration; the naming conventions for the exercises produced in those years are the capitalised terms Year 1, Year 2 and Year 4.

The respondents in the B cohort are also classed as adults; they are university students with L1 G. Because only a cross-section of data was collected from them, no particular naming conventions for their texts were necessary.

The written production of student A7, who was described in Section 1.2, is considered in relation not only to that of the A cohort but also to the B cohort, given her position as both learner and user as defined in Section 1.2. Further information on the content of the texts contained in both the A corpus and the B corpus appears in Chapter 5 of this thesis.

It is generally accepted that formal instruction in the UK has NS-likeness as the FL target norm and this norm is the benchmark for testing purposes.[8] The grammatical structures which school pupils are expected to know are prescribed by the examination boards. At the university which the A cohort attended, grammar instruction is delivered in the early stages with the support of an explanatory grammar textbook (in the case of German, for example, Durrell 1996, Duden 1998 or Dreyer & Schmitt 2001).[9] At this level, too, the criteria for assessment require accuracy of form and correct usage. For the purposes of the present study, the hypothesis which emerges from these statements expects that learners will be aiming to reach NS norms in their assessed performance. In the SLA literature, however, the notion of using NS likeness as a benchmark against which to measure achievement is much discussed, as will be seen in Chapter 2 and which is mentioned in the next section.

1.5   The Second Language Acquisition Context

Having established the linguistic and pedagogical foci of this study and defined the participants from whom the data are elicited, its theoretical underpinnings are introduced.

---

[8] NS-likeness is also known in generative terms as the 'end state'. See Chapter 2.
[9] In Chapter 3, grammatical descriptions of the passive are based on Duden (1998).

Given that preliminary observations in this study noted variation in the production of the German passive not only from learner to learner but also within the production of the same individual depending on the exercise which was set, examining learner language should enable this investigation to contribute to the discussion of developmental sequences in respect of that construction. Support for this contention can be found in Myles' call for investigations of this type: 'If we are to make generalisations about learner development, we need to be able to capture the various stages learners go through' (Myles 2005: 374). Interestingly, much of the initial work in this area was carried out in respect of German and theoretical applications have subsequently been extended to include other languages (e.g. Meisel, Clahsen & Pienemann 1981; Pienemann 1989; 1998); their relevance in this study will be discussed in Chapter 2, as will more recent research which offers a framework to describe the process of acquisition, such as Sharwood Smith & Truscott (2005).

Research in the generative, nativist tradition describes language acquisition (of both first and second languages) in terms of access to Universal Grammar (UG). Lightbown & Spada (2011: 38) suggest that 'while UG provides a plausible explanation for first language acquisition, something else is required for second language acquisition since it so often falls short of full success'. Some SLA discussions have focused on the notion of this 'full success' i.e. NS-likeness (e.g. Bley-Vroman 1983) and maintain that learners' interlanguage should be regarded as an entity in itself and not as an incomplete version of NS language, since it is not clear exactly what native-speaker likeness actually is (e.g. Sorace 2000); others, such as Long (1996) discuss the role of the learning environment and its effect on the learner's activity, which together are said to affect the development of language learning. The present investigation draws on these and similar discussions, and examines interlanguage data collected in a specific formal learning environment. The method of collecting the data is described in the next section.

1.6   The Methodological Context

The first learner corpora were primarily designed to collect, electronically, written data in English produced by NSs of other languages to be used as a support for the teaching of English as L2. These corpora were intended to serve both the instructors who would be able to identify common performance errors and therefore be better prepared to tailor

instruction accordingly, and the learners who would similarly identify errors in the work of others and, after reflection, assess their and others' performance before attempting to restructure their own. Since Pravec identified 10 such learner corpora in English (Pravec 2002: 82-83), collections have grown exponentially to include corpora in languages other than English, both written and spoken, developed by various groups with particular agendas, some of which will be discussed in Chapter 4.3 and 4.4. One of the key benefits of electronic data collections is their accessibility: data, once gathered, can be made available to other researchers for different research purposes, or for verification of claims made by the studies.

As will be shown in Section 1.7 where the research aims are introduced, the intention in this study is to examine evidence from a learner corpus, using it not as described earlier, that is, not to facilitate learning directly, but in an approach which will explore how data collected in this way may be useful in a study of the linguistic, pedagogical and theoretical contexts set out in Sections 1.3 - 1.5.

The use of qualitative analysis is no longer solely associated with social science research in the same way as quantitative research methods are no longer only the province of natural scientists. Both genres draw on quantitative and qualitative methods. In the field of applied linguistics,[10] quantitative studies may contribute to the representativeness of investigations which test theoretical contributions and electronic data collections promote the replicability of conclusions. In SLA studies, a qualitative, case-study-based approach is often chosen (e.g. Bruhn de Garavito 2010). The intention in the study reported here is to combine a quantitative, technology-based analysis with a qualitative, judgement-based commentary in order to offer a contribution to SLA studies.

1.7   The Research Aims and Research Questions

On the basis of the contexts described in Sections 1.3 – 1.6, research aims were established and research questions were framed. The general aim was to investigate patterns in the usage of the German passive as demonstrated by a cohort of learners with L1 English at different stages of learning in the formal environment which was

---

[10] In this study, the term 'applied linguistics' refers to the application of linguistic theory to language pedagogy and to theories of language acquisition.

described in Section 1.2 and subsequently to suggest possible reasons for these patterns. The more specific research aims and the four research questions which were subsequently formulated are set out below.

Research aim 1:

> To contribute to the theoretical discussion concerning the relationship between SLA and second language pedagogy.

Research aim 2:

> To construct an embryonic learner corpus and offer data to a wider research population.

Research questions:

1. Which forms of the German passive are produced in written interlanguage by adult learners with L1 E?
2. Do these forms vary over time?
3. Which forms of the passive are evident in the written production of adult NSs of German?
4. What similarities or differences can be observed between the use of the passive by NSs and NNSs in the respective subcorpora?

The cross-disciplinary study which emerges from these aims applies linguistic theory to that of language pedagogy and language acquisition. By using corpus methodology to provide the data for this empirical investigation, it is possible to analyse the production of multiple learners in a number of learning situations and to observe specific patterns in the production of the passive by learners of German whose first language is English. The description and results of the analytical process will provide the responses to research questions 1 - 4 in pursuit of research aims 1 and 2.

1.8   Naming and Formatting Conventions

Tables and Figures throughout the study are numbered chronologically in each separate chapter e.g. Table 1.1, Table 1.2, Table 2.1 etc., Figure 1.1, Figure 1.2, Figure 2.1 etc.

An asterisk * preceding an example indicates a grammatically unacceptable clause or sentence.

1.9   Concluding Remarks on Chapter 1

Following a general introduction to the early observations which led to the formalising of this study, the cross-disciplinary foundations of the research have been introduced and explained in Sections 1.3, 1.4 and 1.5 of Chapter 1. An introduction to learner corpora was the subject of Section 1.6 and 1.7 contained the research aims and specific research questions which emerged from the preceding considerations.

Looking ahead, Chapter 2 of the thesis contains a review of some of the relevant literature in the disciplines of linguistics, language pedagogy and SLA in order to situate the study against its theoretical background. In the same chapter, the development of learner corpora is set within the historical context of corpus linguistics, and the relatively new contribution of such corpora to research into the acquisition of second languages other than English is pointed out. Chapter 3 presents a detailed theoretical description of passive syntax and its varied semantic interpretation in German and the application of that theory in teaching practice in a formal learning environment in the UK, such as the one in which this study was carried out. Chapter 4 brings together the corpus methodology outlined in Chapter 2 and the description of the passive in Chapter 3 to focus specifically on learner corpora in German. A comparison of existing corpora of this type indicates the body of research within which the corpus described in the present study should be located. A number of alternative analytical systems are then examined, culminating in the choice of the mixed-method analysis to be used here. Finally, the steps are described that were taken to collect the data for this study prior to analysis, including the design of the corpus. The quantitative data extracted from it are analysed in Chapter 5 and the results of a type of qualitative analysis of the entire corpus are reported in Chapter 6 in a series of mini case studies, where the acquisition of the syntax of the German passive is assessed in relation to its

meaning in the learners' chosen context. A more in-depth examination of the passives produced by the NNSs in their final year of study is presented in Chapter 7, comparing the learners' production of the passive with that of the NSs in the data sample. These analyses are followed in Chapter 8 by an evaluation of the findings in the light of theoretical considerations, together with an assessment of the relative achievement of the research aims and quality of response to the research questions. A further section of that chapter reviews the place of this study in its research context and takes a look ahead towards possible routes of future enquiry, based on issues that have emerged in the course of the investigation.

The Appendices A – E comprise documents which were devised by the author in pursuit of the research aims.

The contents of the corpus are described in Chapter 4.6. Copyright rests with the author of the thesis to whom all requests for access should be addressed.

CHAPTER 2 - LINGUISTICS, SECOND LANGUAGE ACQUISITION AND LEARNER CORPORA: THE INTERDISCIPLINARY RESEARCH CONTEXT

2.1 Introduction

The reviews which are contained in this chapter consider a number of studies with different research agendas all of which provide the foundations for the study described here. One point of view, for example, can trace a link back to a nineteenth-century German linguist, the co-founder of the University of Berlin, Wilhelm von Humboldt, whose opinion that 'we cannot really teach language: we can only present the conditions under which it will develop spontaneously in the mind in its own way' was published in 1836 and is quoted in translation by Chomsky (1965: 51) and in an almost identical translation by Corder (1967: 12). This statement will find favour with anyone who supports the contention that teaching has little direct effect on language development. Another point of view is represented by exponents of specific, explicit instructional methods such as Long and Robinson (1998). In their paradigm, distinctions are made between language forms and language functions; in the study described here both are important since the two syntactic forms on which it focuses each has a different semantic function. In accordance with the first stated research aim, to contribute to the theoretical discussion concerning the relationship between SLA and second language pedagogy, the effect of instruction will be considered in this thesis in a framework which considers theories from more than one theoretical stance.

The diversity of forms used to express the passive across languages has offered a wide-ranging and dynamic spectrum for researchers: the construction has been considered as a syntactic form and its semantic meaning has been discussed, both by using corpus-based methodology (e.g. Granger-Legrand 1976 for French and English) or focusing on the study of a single case (e.g. Abbot-Smith 2003 for German). These types of studies, together with those which have looked more generally at language acquisition both by children and by adult learners, provide a framework for the analysis of L2 learner language which is described here.

This second chapter is constructed in the following way: linguistic aspects of the acquisition of the German passive are discussed, followed by a review of a number of theoretical considerations of instruction in SLA. The final section places the field of learner corpora in its historical context and explains why this particular method of collecting data has been chosen for this study.

Firstly, a brief definition of the passive with cross-linguistic validity is offered[1] and the term 'periphrastic passive' is identified, which is applicable to a number of theories of grammar constructs, as is the term 'basic' passive. Since the vast majority of studies of the acquisition of the construction focus on English both as L1 and L2, some of those are reviewed because they provide useful linguistic reference points for this study. Studies of the acquisition of the same construction in German are then reviewed; these are few in number, as far as it has been possible to ascertain. In Chapter 2.3, two studies of second language acquisition at tertiary education level are reviewed in order to indicate the type of approach used in studies similar to the one described here.

Thereafter, further reference points are established with reference to SLA research in the generative and functionalist traditions, both of a general nature and that which is more specifically related to second language learning and teaching. Interestingly, one strand of this type of research has focused on German since the early days of SLA theory in the 1970s, as pointed out in Chapter 1.5, but the acquisition of the passive construction in that language has rarely been discussed.

The final section of the chapter examines the relatively new genre of learner corpora, indicating how it has developed from its initial focus on evidence from corpora in L2 English to include corpora from learners of other languages such as the LINCS corpus described in this thesis. This second chapter concludes with a description of some of the extant corpora with which the LINCS corpus might be compared or to which it might be offered in due course, in resolution of research aim 2:

To construct an embryonic learner corpus and offer data to a wider research population.

---

[1] An additional account of the passive terminology which is used more specifically in a pedagogical context appears in Chapter 3.

The way in which the treatment of these corpora influences the methodology used in this study is explained in Chapter 4.

2.2   The passive

2.2.1   Passive Taxonomy and Typology

In general terms, both active and passive grammatical forms allow the listener/reader to hear/view the action or experience expressed by the verb in a different way, depending on the focus conveyed. Sentence structure and/or grammatical variation will indicate where attention is directed.

Durrell's explanation of the perspective offered by the passive serves as a useful introduction. 'Using the passive allows us to talk about an activity without mentioning who is doing it' (Durrell 1996: 294). This applies not only to natural languages but also, according to Moskovsky and Libert (2004), to possible passives in an invented, artificial language, in which, they indicate, passives should be included as they offer the only way to express the particular function mentioned by Durrell.

Grammatical morphology of one type or another is usually necessary to express the passive, (with the exception, of course, of languages where the construction does not exist, such as some in Indonesia and on the west coast of South America), and the expression may involve a choice between many forms such as in Malagasy (Keenan 1968: 279). In the case of German, two alternative syntactic forms of the passive construction are formed by combining an auxiliary verb (AUX) with a past participle (PP) as already indicated in Chapter 1.3.[2] In English, a total of seven different types of 'be' + PP constructions are identified by Granger (Granger 1993, quoted in Granger 1997). As the study described here focuses on the acquisition of German by L1 English speakers, it will refer particularly to the means of expressing a passive in those two languages, indicating the similarity and dissimilarity between them in the course of Chapters 2 and 3.[3] Both in generative and in functional terms, a change in the sentential position and function of the grammatical subject/topic and object/focus is necessary to express a passive, where the meaning shifts from the grammatical subject of the active

---

[2] A detailed explanation appears in Chapter 3.3.1.
[3] Similarities and differences between English and German were alluded to in Chapter 1.3.

sentence doing an action to experiencing or feeling that action or situation in the passive alternation.

The various definitions which follow in this section introduce the terminology which is common to studies of the acquisition of the passive and which refines the area of interest in this particular one.

Keenan (1985) considers the verb form as the determining element in a sentence, which may or may not subsequently cause the alteration of other syntactic elements. In the examples which follow in Chapter 3, the aim will be to show that the passive in German must be examined from both a syntactic and a semantic perspective.

The grammatical structure of the passive in German fits the taxonomy described by Keenan (1985), which may be summarised as follows:

- within a passive clause, the direct object of the verb in the corresponding active clause becomes the subject of the passive verb;
- the subject of the active verb may be omitted in a passive clause; if it is present, it will be represented by a prepositional phrase or will be marked for case;
- the passive verb will show passive verb morphology.

Additionally, Keenan (1985: 247) divides passive forms into 'basic' and 'complex': basic passives are described as short sentences containing only a noun phrase (NP) which at this stage of the study is used to mean a determiner plus noun ('the wine' in (1), below) and a verb phrase (VP) ('was drunk' in (1), below), with the optional agent ('the scientist' in (2), below) being omitted and the verb being transitive and expressing an activity.

(1) The wine was drunk.

In comparison, we may note the difference between a basic passive according to this definition and St John and Gernsbacher's 'simple' passives (St John & Gernsbacher 1998: 240), which are typically of seven words in length and include a 'by' phrase as shown in (2).

(2) The governor was advised by the scientist.

Keenan's term 'complex' passives refers to verbs with complex internal arguments such as phrasal verbs (e.g. to give up) and ditransitive verbs (e.g. to give, to bring). In the context of German, complexity will refer to verbs which are compulsorily followed by a case other than the accusative. Such verbs are identified in Chapter 3.3.1.1. Keenan & Dryer's general point that '[l]anguages with basic passives commonly have more than one formally distinct passive construction' (Keenan & Dryer 2006: 340) also applies to the study of the German passive with its two specific syntactic forms. The study reported here will concern basic passives as defined by Keenan although reference will be made in Chapter 3.3.3 to the additional 'by' phrases included in St John & Gernsbacher's (1998) simple passives.

Basic passives are further subdivided by Keenan into strict morphological passives, where the passive is formed by the addition of a verbal morpheme, and periphrastic passives which are formed with an auxiliary verb together with a morphologically distinct transitive verb. In this study, we are not concerned with the NP as described above; the focus here is on the VP as defined by Greenbaum & Quirk (1995) and which is discussed more fully in Chapter 3.2.1. This VP is composed of two parts, AUX plus PP, as mentioned above, and to that extent conforms to the term 'periphrastic passive'. As will be apparent in Chapter 3, however, constructing the passive in German is not limited to transitive verbs.

In addition to the terms already mentioned, there have been some attempts to refine the description of passive typology still further (e.g. for English and Serbian (Djurkovic 2004) and for English and Arabic (El-Marzouk 1998) into passives and impersonals. The latter type also appears in German, either without a subject or with the pronoun *es* (= it) in the first position as in (4).

Keenan's description of a periphrastic passive is borne out by German grammar handbooks such as Duden (1998)[4] and Durrell (1996).[5] The description holds true both

---

[4] Duden volumes encompassing descriptions of and examples from all aspects of the German language, from orthography (volume 1) to quotations (volume 12) are the standard prescribed reference works for the German language in Germany. See also Chapter 3.

[5] Durrell's pedagogical grammar of German is written by a professor in the tertiary education sector in the UK and as such is intended for students such as the A cohort in this study.

in purely grammatical discussions of the types of verbs which can and cannot passivize (Bresnan & Kaplan 1982) and in functional linguistic terms of the relative weight of the promotion of the object versus the demotion of the subject (examined in Djurkovic 2004).

2.2.1.1 Salience

Djurkovic's interpretation would only hold true in German in the case of simple passives where, in semantic terms, both agent and patient are present. In the functionalist approach as illustrated by Djurkovic (2004), items which are placed either in first or last sentential position are deemed most salient, occupying 'peaks of prominence' (Halliday & Matthiessen 2004: 61). This functionalist notion has a parallel in Slobin's (1985) description of operating principles in L1 acquisition whereby children pay most attention to the beginnings and ends of sentences. However, it should be noted that in the case of a German sentence or clause where an adverbial rather than an NP may occupy the first position, and where the conventions of German word order dictate that the PP occurs in the final position, neither the NP nor the agent 'by'-phrase is foregrounded (3) and this results in the first and therefore the salient item in (4) giving no indication of meaning.

(3)   Am Abend   wurde   viel Wein   von   den   Gästen
      getrunken.
      In the evening became/was becoming   a lot of wine   by the guests
      drunk.
      A lot of wine was drunk by the guests in the evening.

(4)   Es   wurde   während des Oktoberfests   in den Zelten
      viel   getrunken.
      It   became/was becoming   during the Oktoberfest   in the tents
      much   drunk.
      People were drinking/drank a great deal in the tents during the Oktoberfest.

In terms of processing, it is not possible to interpret the sentence until the final verbal element in (3) and (4) has been uttered. For example, the PP *gelacht* (laughed) could be substituted for *getrunken* in (4).

2.2.1.2 The Passive as a Construction

The term 'construction' should be clarified since it is applied differently according to theoretical orientations. Offering a purely grammatical description, Chalker & Weiner suggest that the passive is a construction 'involving a passive verb' (1994: 285), in other words, a verb showing passive morphology. The implication of this statement is that the verb will be an integral part of an utterance which will convey passive meaning and which will in some way be a combination of items. This accords with Keenan's (1985) view, as set out in Section 2.2.1. In constructionist theoretical terms, however, constructions are the building blocks of language learning and are 'understood to be learned on the basis of positive input' (Goldberg 2003: 22). According to this theory, language develops in the same way as other cognitive processes, through repetition and practice, in this case through 'repeated use' of constructions (Ellis 2007: 78), which can be as small as an added morpheme e.g. –ing or as large as a passive sentence following the order Subject Verb Object (Whong 2011: 76). 'Constructions are a product of language use and not the product of grammar rules', explains Whong (2011: 76). In contrast, according to the generative view, language develops according to the principles defined by Universal Grammar (UG) within the language module in the brain, principles which limit what is possible in terms of the development of a learner's grammar. In both theoretical frameworks, exposure to multiple examples of language is necessary, but generativists point out that child language acquisition does not occur simply on the basis of positive evidence in input which cannot provide every instance of possible alternations. How does a learner identify whether the lack of an item in the input, sometimes known as indirect negative evidence, is due to 'coincidence' (Gass, Behney & Plonsky 2013: 163) or the fact that it is disallowed in the grammar? According to the generative view, if children are able to produce grammatical items which they never hear in the environment, innate principles must exist which constrain acquisition.[6]

If the term 'construction' is used in this study, it will be used in the traditional, non-theoretical sense that the passive in German is 'constructed' i.e. its form is produced by

---

[6] 'Negative evidence' is the term used to refer to what is not permissible in a language grammar and may be expressed explicitly (as in correction) or implicitly through repetition or a recast. Explicit corrective feedback is rarely given to children by carers or parents and children are known to pay little or no attention to it (Pinker, Lebeaux & Frost 1987: 196). Recasts are not guaranteed to be recognised as corrections (Gass 2003). In a second language learning classroom situation the term refers to a kind of explicit instruction about what is not possible in a given language.

combining two or more grammatical elements, AUX and PP. This interpretation would seem to be applicable to more than one theoretical stance.

2.2.2   Studies of the Acquisition of the Passive

Having introduced the theoretical descriptions of the passive which will be used in this study, the next section draws on studies of the acquisition of the passive both in English and in German. Thereafter, two studies are considered which focus on adult learners of second languages, similar to those in the A cohort described here, as indications of the varied approaches to examining language acquisition which have been adopted.   In Section 2.3, one of these concerns the development of L2 German and draws its data from a similar student cohort to the one described in the present study, looking at the passive among other structures, and the second concerns the acquisition of the passive in L2 English. At the end of the section a preliminary hypothesis is proposed regarding the developmental sequence of the two passive types which are the focus of this study.

2.2.2.1 Studies of the Passive in English L1

From the generative perspective, Borer & Wexler examined 'verbal' and 'adjectival' passives (Borer & Wexler 1987: 141) in English and in Hebrew and concluded that only adjectival passives are present in early child grammars. In the study described here, 'verbal' passives equate to passives constructed with the auxiliary *werden* and 'adjectival' passives to those formed with *sein.* These are the two basic syntactic forms of the German passive which are described in more detail in Chapter 3.3.1. Borer & Wexler (1987: 141) term these verbal passives 'short' or 'truncated' passives where a passive participle is used adjectivally; according to their assessment they are formed only from action verbs and are never accompanied by a 'by' phrase (5).

(5) The island was uninhabited (Borer & Wexler 1987: 141).

In contrast, in German, so-called adjectival passives can indeed be accompanied by a 'by' phrase, and Pinker, Lebeaux & Frost point out that Borer & Wexler's contention was contradicted by Wasow (cited in Pinker, Lebeaux & Frost 1987) with the example 'This island is uninhabited by humans'.

In Borer & Wexler's terms (Borer & Wexler 1987) those constructions which are accompanied by a 'by' phrase are denoted as 'full' passives (6).

(6) Your letter of thanks was much appreciated by the whole family.

Taking their evidence from the L1 acquisition of the passive in English and in Hebrew, the reason which Borer & Wexler advance for the primacy of emergence of adjectival passives over verbal ones is that a child's mental grammar matures and certain structures will not be produced until that grammar has reached the maturational point where it can assimilate verbal argument structure. This hypothesis is followed by those who accept the existence of innate language structures (the properties of a universal grammar) but who contend that these structures are not all available to the infant at birth; some of them take time to be used systematically.[7] Weinberg's comment on Borer & Wexler's article (Weinberg 1978) concludes that their findings do not only provide support for the Maturational Hypothesis; she points out that 'non-actional passives are rare in the child's early linguistic input' (1978: 178) and that this may equally well account for the fact that they occur rarely in the output. This view is supported by Pinker, Lebeaux & Frost (1987: 200) who say that '[p]assives of all sorts are rare in parental speech'. If this is the case, how do children come to be able to produce passives when the item in the target language input is relatively infrequent? Followers of the Maturational Hypothesis such as Borer & Wexler maintain that acquisition depends on the gradual development or maturation of features of a child's grammar. On the other hand, according to those who support the Continuity Hypothesis, such as Pinker (1996), the basic grammatical structures of a universal grammar are present and develop as the child learns the semantic roles and thematic functions of, for example, certain verbs from the input. Arguing from the perspective of Lexical Functional Grammar (LFG), he explains that the passive is encoded by the generation of the lexical entry for the active verb.

The following example from Pinker (1996), expressed in generative grammatical terms, indicates the complexity of mechanisms involved in producing a passive. The lexical entry in a child's grammar of the verb 'to eat', for example, will propose an object (e.g.

---

[7] The discussion of the 'developmental' versus the 'logical' problem is neatly summarised by Cook (Cook 1993: 209): 'Development is concerned with the chronology of emergence; acquisition pays no attention to time.'

an apple); the grammar will change this object to subject (OBJ>SUBJ), then convert the subject (e.g. the horse) to agent or oblique (SUBJ>OBL) and finally activate the verb morphology (MORPH) which in the case of the passive is the past participle, giving the sentence *The apple was eaten by the horse* (optional OBL). In practice, Pinker says, to form this kind of sentence children must rely on meaningful, contextually sound input from adults to form grammatical rules; only after the stimulus of input can the process of production begin.

In their earlier study, Pinker, Lebeaux & Frost (1987) had treated the learnability problem in respect of the passive in English in a discussion of thematic relations and the ability of a child's grammar to represent certain activities. They concluded that 'there is no reason to believe that young children's passives are exclusively adjectival' (1987: 263), particularly as most parents use action verbs when describing events to their children, thereby expressing a view contrary to that of Borer & Wexler (1987).

Evidence from Slobin's (1981) report on child spoken language across languages suggests that children's doubts as to the agent/patient relationship might prove problematic for an understanding of the passive. When considering L2 acquisition by adult learners, however, it must be remembered that they will not only be familiar with particular constructions and their associated concepts in their L1, but in a formal learning situation they will also be receiving instruction in the linguistic terminology with which to explain them. It is worth noting that the interactionist perspective on L2 learning taken by Mackey, Abbuhl & Gass (2012) points to the positive value of interaction of all kinds, whether with teachers, peers or native speakers, in enabling a learner to make progress.   With second language learners in a formal learning environment having the benefit not only of additional conceptual knowledge but also supplementary grammar scaffolding from instructors, it may be that it is the non-existence of specific items in the L1 (in the present context, an alternative passive auxiliary) which causes problems of acquisition for the L2 learner, as suggested in Chapter 1.3.

2.2.2.2 Studies of the Acquisition of the German Passive

While the English terms 'verbal' and 'adjectival' are used by Borer & Wexler (1987: 141) to differentiate between two passive types as has been mentioned in Section

2.2.2.1, these are not the only terms to be chosen in this context. Abbot-Smith (2003) and Bruhn de Garavito (2010) opt for 'eventive' and 'stative'. The expressions used in grammar handbooks are identified as 'dynamic' (Greenbaum & Quirk 1990: 45) or 'actional' (Chalker & Weiner 1998: 8) to describe the process of the action and 'statal' (both volumes) to describe the finished state. In this study 'actional' is favoured since it is specifically named as the contrast to 'statal' by Chalker & Weiner (1998: 372), 'stative' being applied to denote word classes such as verbs or nouns, whereas 'statal' is the descriptor for a passive type (Chalker & Weiner 1998: 371). Durrell's grammar handbook for students with L1 English differentiates between '*werden*-passives and *sein*-passives' (Durrell 1996: 294). In the limited amount of literature describing studies of the passive written in German, Bryan (1995) favours *Vorgangspassiv* and *Zustandspassiv,* which are the terms used by the German grammar handbook Duden (1998: 183). The designations '*werden* passive' and '*sein* passive' will denote the two types of passive which are examined in this study. These equate to the terms 'actional passive' and 'statal passive' respectively, which will also be used.

Both of these types can be traced back historically to Indo-Germanic forms which occurred in Gothic, namely the *waírþan* - Passiv, and the verb *wisan* + adjective which may be considered as the forerunners of the modern German *werden* and *sein* passives respectively (Lühr 2010: 171). The designation of one of the Gothic forms as verb + adjective is interesting for teachers since it suggests that this might be a preferred way to describe the construction to learners as a way of differentiating between the potentially confusing two types of passive. This suggestion will be considered in Chapter 8.

As has been indicated in Section 2.2, contributions to the study of the acquisition of the passive in L1 English have come from different theoretical backgrounds e.g. syntax (Keenan 1985) and lexical theory (Pinker, Lebeaux & Frost 1987). It should also be noted that the number of SLA studies of the English passive is legion (see Wang 2010 for a recent overview). It can be seen from Tables 2.1 and 2.2, however, that in SLA the construction in German has received no specific attention as far as has been ascertained, studies of German interlanguage tending to focus on word order or morphology (e.g. Schwartz & Sprouse 1994). A more recent paper which reviews a spectrum of research covering the same disciplines as the ones described in this thesis refers in general terms to verb morphology but the passive is not mentioned (Eckerth,

Schramm & Tschirner: 2009: 47). The studies in Table 2.1 are based on elicitation tasks and grammaticality judgement exercises, such as the example in Appendix B.[8] Table 2.2, on the other hand, displays the two examples of corpus-based studies of the German passive that we have identified, both of which are based on the same data sample.[9] It is possible that the complexities of the construction which are described in Chapter 3 may account for the small number of studies which have addressed its developmental patterns. The present study intends to add to that number.

The three L1 German studies, Eisenbeiß (1993), Bryan (1995) and Abbot-Smith (2003) will provide useful comparative evidence for this study for the following reasons: Abbot-Smith's study is corpus-based but founded on construction theory (i.e. usage-based) as described earlier in Section 2.2.1.2; Bryan's and Eisenbeiß's are founded on generative theory but their methodological approach is experimental. In the present study, the corpus-based methodology chosen by Abbot-Smith is discussed in terms of the generative theoretical principles favoured by Bryan and Eisenbeiß, a combination which has hitherto received little treatment. Table 2.1 lists a number of experimental studies of the passive construction, compared to just two which are supported by evidence from a corpus dataset as shown in Table 2.2; it is intended that this study will redress that balance. Although the two studies in Table 2.2 are corpus-based in that they investigate longitudinally-collected data, they are based on examples from a single L1 case and are therefore not classed as learner corpora as defined in Section 2.6.3.1. In the study which is the subject of this thesis, the corpus comprises data collected from a number of learners rather than just one subject, and the study examines L2 acquisition rather than L1. This combination of analytical measures is not seen in any of the studies in Tables 2.1 and 2.2.

---

[8] This grammaticality judgement exercise (GJE) was administered to the A cohort of respondents. A preliminary analysis has been carried out but the results are not presented in this thesis.
[9] The same data have also been discussed in Abbot Smith & Behrens (2006).

Author/year	Title	L1	L2
Bruhn de Garavito, J. (2010)	Eventive and stative passives: the role of transfer in the acquisition of *ser* and *estar* by German and English L1 speakers, in Collentine, J., García, M., Lafford, B. and Marín, F.M. (eds.) *Selected Proceedings of the 11th Hispanic Linguistics Symposium*. Somerville, MA: Cascadilla Proceedings Project, 27-38.	E/G	S
Bryan, A. (1995)	*Der Erwerb des Passivs: sprach- und lernbarkeitstheorethische Betrachtungen zum Englischen und Deutschen*. PhD thesis, Universität Tübingen.	G	
Eisenbeiß, S. (1993)	'Auxiliaries and the acquisition of the passive', *Proceedings of the 25th Annual Child Language Research Forum*. Stanford, CA: 235-242.	G	
Flett, S. (2006)	*A comparison of syntactic representation and processing in first and second language production*. PhD thesis, University of Edinburgh.	S	S
Hikima, N. (2010)	*The effects of processing instruction and re-exposure on interpretation discourse level tasks: the case of Japanese passive forms*. PhD thesis, University of Greenwich.	E	J
Montrul, S. (1999)	Causative errors with unaccusative verbs in L2 Spanish. *Second Language Research*, 15, 191-219.	T/E	S
Yip, V. (1995)	*Interlanguage and learnability from Chinese to English.* Amsterdam: John Benjamins.	C (Taiwan)	E

Table 2.1: Examples of experimental studies of the acquisition of the passive

(C = Chinese; E = English; G = German; J = Japanese; S = Spanish; T = Turkish)

Author & year	Title	L1	L2
Abbot-Smith, K. (2003)	*Piecemeal paths to grammatical productivity: how children become productive with basic event constructions*. PhD thesis, University of Manchester.	G	
Abbot-Smith, K. & Behrens, H. (2006)	'How known constructions influence the acquisition of other constructions: the German passive and future constructions'. *Cognitive Science*, 30, 995-1026.	G	

Table 2.2: Corpus studies of the acquisition of the German passive

G = German

Mention has already been made of the small number of studies which have focused on the German passive, despite a considerable amount of interest being directed towards the passive in English. This is the case both for German L1 and German L2. There is therefore little evidence from which to draw comparisons for the purposes of this study. For that reason, we propose to examine more closely two of the studies listed in Table 2.1 which are based on generative language acquisition theory and a third from Table 2.2 which is rooted in construction theory. In continuing the investigation of the relationship between SLA and second language pedagogy, identified as research aim 1, the intention will be to explain how previous studies of the acquisition of the German passive lead us to propose a particular acquisition order of the various ways of expressing it by adults with L1 E learning German as L2. In Chapter 2.5.3.3, where SLA theoretical accounts of developmental sequences are discussed, it will be seen that development follows a predictable natural order which seems to remain unchanged despite teaching input. The aim of this study is to confirm or disconfirm whether the development of the two syntactic forms of the passive in L2 adult acquisition of German follows the same sequence as in the L1. All of the three studies discussed in this section which focus specifically on the acquisition of the passive relate to German L1 (Abbot-Smith 2003, Bryan 1995, and Eisenbeiß 1993); following the data analysis in this study, it will be possible to determine whether the findings are replicated in the investigation of the acquisition of German L2. Initially, a review of a more broad-based study of the acquisition of German is included (Mills 1985) which features the acquisition of the passive in one of its subsections. Once theoretical considerations have been addressed in the light of the findings, a preliminary hypothesis is proposed in Section 2.4 which will be tested by analysing the learner corpus.

The first study under consideration is that of Mills (1985). Mills concludes that passives appear in German sporadically in child language up to the age of four and are 'always agentless'. She further describes the passive as a 'complex structure' (Mills 1985: 156) which appears later than simple ones and which 'is used quite rarely by children' (Mills 1985: 201). Her finding that *sein* passives are seen to appear before *werden* passives mirrors Borer & Wexler's claim that adjectival passives are present in children's grammars before verbal ones. Mills also notes that use of the auxiliary *werden* (which in German is not only the required auxiliary for the formation of action passives but also for the formation of the future tense and in addition functions as a copula) appears with both passive and future structures at approximately the age of three.

Eisenbeiß (1993) points out that it is not only the auxiliary *werden* which has more than one grammatical function: *sein* similarly not only acts as a passive auxiliary but also as a copula verb and as the auxiliary for a sub-category of verbs in the construction of the perfect tense. Eisenbeiß (1993) carried out an experimental picture study of the development of *werden* passives and *sein* passives in relation to their dual (*sein*) or triple (*werden*) functions, namely the construction of the future tense (*werden*) and use as a copula (both *werden* and *sein*). Her results show that both types of passive are produced by young children and that there is a correlation between their correct interpretation of passive and copula constructions and their use of the same verbs in passive utterances. Eisenbeiß therefore claims that her findings are in opposition to Borer and Wexler's (1987) maturation-based contention because both types of passive are produced concurrently from the age of two. On the other hand, her findings support a lexicalist theoretical point of view (e.g. Pinker, Lebeaux & Frost 1987) that acquisition of the passive occurs as a result of a child's grammar accommodating the lexical properties of the construction, supported by learning of other, similar constructions such as, in the case of the German passive, copula verbs.

Countering Eisenbeiß's results, Bryan (1995) showed in her experimental study of two German children and the development of their L1 that *sein* passives are acquired before *werden* passives, even though a *werden* passive appeared in the language of both children before a *sein* passive. In both cases, after the first use of a *werden* passive the form disappeared for a time and reappeared only after the *sein* passive had been acquired (Bryan 1995: 214). The meaning of 'acquired' here is that use of the form

became systematic. A more detailed discussion of the meaning of 'acquired' follows in Section 2.5.3.5.

The contrast in the results of the last two studies mentioned led Abbot-Smith to conduct a further study of the acquisition of the *werden* and *sein* passives (Abbot-Smith 2003). Her results accorded with those of Bryan in terms of the acquisition orders, indicating that *sein* is acquired before *werden*. She concludes that 'children will become productive with a grammatical construction more quickly if they have previously learned simpler related constructions' (Abbot-Smith 2003: 138), indicating that acquisition of a construction does indeed depend on whether the features of that construction, evident in other morphological or semantic forms, have already been successfully acquired. In this, Abbot-Smith's claims accord with those of Eisenbeiß.

2.3   Studies of Language Acquisition by Advanced Learners

In general, SLA studies of various types from a number of theoretical perspectives[10] have investigated learner language using data elicited experimentally, and often focusing on spoken discourse, and to a lesser extent on written data. As will be indicated in Chapter 2.6.3.2, studies based on learner corpora are fewer in number because of the relatively young age of the sector. The collection of papers in Gilquin, Papp & Díez-Bedmar (2008), however, contains examples of studies more similar methodologically to the one described here, drawing on written learner corpora and using CIA methodology as described in this Chapter (Section 2.6.2).

More specifically, having focused in the previous section on the few studies of the acquisition of the German passive which it has been possible to identify (none of them in L2), two studies of the acquisition of the passive by advanced L2 learners are now considered. In terms of learner type, these match those in the present study. The studies carried out by Klapper & Rees (2003) and Hinkel (2004) examine the effects of language instruction exemplified by students' written production of the passive in German and English respectively.

Klapper & Rees (2003) assess the progress made by their cohort of students of German in a longitudinal study looking at the improvements in performance after their period of

---

[10] See Gass & Mackey (2012) for a comprehensive overview.

study abroad of three groups of students who had undergone varying degrees of form-focused instruction (FFI). [11] The students' production of the passive improved, as did their target language generally, but the results did not indicate whether benefit was derived from one pedagogical methodology over another, or from the period of residence abroad.

Hinkel's (2004) study looks at the acquisition of the passive specifically in English L2 and differs from those looking at L1 acquisition previously mentioned in Section 2.2.2.1 in that it is focused on academic written discourse in which she indicates that 'usage of the passive voice has (…) been commonly associated with formal prose across diverse types of writing' (Hinkel 2004: 23). While Hinkel's focus is on L2 English, it will be of interest in this study to discover whether there is a similar association in German of the passive with academic discourse and if that is the case, to what extent it appears in the L2 German in the LINCS corpus.

In her evaluation of the quality of learner language, Hinkel (2004) recommends FFI in order to improve the production of passives. She does not find a correlation between frequency of input and L2 learner output. She cites as an example present passives which are 'so common that they are usually familiar to L2 learners even at the beginning or intermediate levels of L2 proficiency' (2004: 25). Despite this, Hinkel finds that learners avoid complex structures such as the passive. An examination of the B corpus in this study will permit some comments to be made relating to the use of complex language by NS users of German as compared to what is demonstrated by a sample of the production of adult learners.

Taking an example from Hinkel's findings, it will be relevant in the present study of the LINCS corpus to consider learners' use of different tenses in the German passive. According to the example from Klapper & Rees, we may expect there to be some improvement in learner production between years 2 and 4 (following a year abroad in year 3), although exact information regarding input during that third year will not be available and will not form part of the evidence under consideration in this study. FFI will, however, be a topic for discussion in Chapter 8 where the effect of pedagogical instruction on language acquisition is discussed.

---

[11] See Section 2.5.3.2.

2.4   Preliminary Hypothesis Generation

The studies of child language acquisition of German reviewed in section 2.2.2.2 show results only in respect of *sein* and *werden* passives. The study reported here also considers the use of alternatives to the passive because they are described in the grammar textbooks as indicated in Chapter 3.2. It is perhaps unsurprising that reports of the acquisition of these alternatives do not seem to have occupied researchers, given the syntactic and semantic manipulations which are needed to produce the basic passive type.

Despite their differing theoretical foundations, all three L1 passive-focused studies (and the broader-based one of Mills 1985) reach the same conclusion: that *sein* passives are acquired at the same time as or before *werden* passives. However, this study set out to investigate whether L2 learners are different, based on evidence from classroom observations, as suggested in Section 2.2.2.1. The study seeks evidence to support the claim that, in adult L2 German, where the learners' L1 is English, the *sein* passive is acquired[12] after the *werden* passive, in spite of the fact that the *sein* passive is closer in grammatical form to the English passive than the *werden* passive.[13];[14]   In Section 2.5.3.9, the notion of transfer from the learners' L1 will be discussed, along with other terms in Section 2.5.3 which will be of use in the description of emerging learner interlanguage.

2.5   Second Language Acquisition and Second Language Pedagogy

A number of criticisms have been levelled at SLA research by FL teachers because its theoretical models appear to have little relevance to language pedagogy. Ellis, for example, suggests that theoretical underpinnings would carry more weight if related to their practical applications expressed in the following way: 'Relevance [of a theory] must necessarily be determined (…) by demonstrating how the findings of SLA address the needs and concerns of practitioners' (Ellis 1997: 32). A different contention is

---

[12] 'Acquired' here means that the item is used in a pragmatically and syntactically correct way i.e. both function and form are correct. A discussion of the ways in which 'acquisition' is interpreted in various studies follows in Section 2.5.3.5.

[13] According to a preliminary analysis of a questionnaire (Appendix A) to which the A cohort responded, the *sein* passive is the form which is perceived as the more difficult one by L1 E learners.

[14] It should be stated, however, that no firm claims are to be made regarding the development of the learners in this study because of the small corpus size and the variation in the text types which are analysed.

offered here: the value of SLA research is as a discipline in its own right; it is up to teaching practitioners to select from that research what is beneficial for their own pedagogical practice.

Another proposal for combining the two disciplines comes from Lakshmanan & Selinker who suggest that SLA research in the pedagogical domain could well be carried out by 'bilingual researchers' who have at one time been FL teachers themselves and therefore have an insight into language development as both learner and facilitator. They would know 'several relevant linguistic systems: the native language of the learner, a very advanced state of the interlanguage, earlier interlanguage systems and the target language itself, at least in a declarative sense.' (Lakshmanan & Selinker 2001: 415). Romaine's definition of bilingual (footnote 4, Chapter 1) would also seem applicable here, in a context where the research has been carried out by a teaching practitioner who has learned a second and subsequent FL only in formal learning environments. Research question 1, as set out in Chapter 1.7, is addressed with these observations concerning the practical application of theory to language teaching in mind.

Myles has also called for closer ties between SLA and FL teaching (Myles 2005) and there is a growing body of literature which situates classroom-based research firmly against the background of SLA theory, not least studies carried out by Myles herself (e.g. Myles 2004) in collaboration with Mitchell and their teams of researchers in Southampton and Newcastle which are discussed further in Section 2.6.3.2. These are examples of the product of empirical research carried out in a formal learning context.

2.5.1 Language Teaching Methodology

Considering that training learners to learn a second language is what teachers of a foreign language do, the following statement from Cook (1996: 1) might seem somewhat anomalous: 'Most teachers have been trained to teach, not to think about second language learning'. The essence of it is, though, that teaching outcomes are prescribed by the educational systems in different countries (Cook 2008: 9; Whong 2011: 23) and teachers are necessarily more concerned with a learner's performance than how it was achieved. In secondary schools, prescribed textbooks adopt certain methodologies and even the study of literature cannot be relied upon to further students'

language skills because of the requirement for pupils to respond to rather broad thematic questions on a set text. The way that learners actually learn the language traditionally plays no part in teaching at this level, which in any case is subject to constraints of time and in some geographical areas to the availability of teaching staff. Considering that it has been suggested that the origins of SLA 'lie in a practical orientation to language teaching' (VanPatten & Williams 2007: 17), it is strange that more cross-fertilisation of ideas has not taken place, either at secondary or tertiary level. This study offers an addition to the literature addressing a direct link between teaching and SLA theory.

The methodologies of language teaching have varied since the 1960s when the audio-lingual method was still popular. Learners were expected to listen to dialogues and respond to them by copying drills and practising pronunciation, often in the first generation language laboratories which represented state-of-the-art technology at the time. It is possible to tie the decrease in popularity of this behaviourist – oriented method to the publications of Chomsky's early linguistic hypotheses concerning the development of language in the brain by way of a specific language module (e.g. Chomsky 1965: 33). Within the human mind, input received by a child activates the features of an innate language system which are represented in the child's environment; for example input in Spanish will lead to a child learning Spanish omitting subject pronouns whereas for a child whose mother tongue is English, input in English will result in subject pronouns being operationalised. Activation of a new structure, according to this view, can be the result of a single occurrence in the input.

In the last four decades, the pedagogical pendulum has swung in the direction of communicative language teaching (CLT), based on the notion that learning language in a functional context will lead to more successful acquisition outcomes. In this communicative paradigm, the amount of time devoted to the teaching of grammar is a subject for debate but is generally regarded as minimal. According to Spada (2007: 272), the definition of CLT depends on whom you ask. By the 1980s, as she points out, the view prevailed that language teaching should be 'exclusively meaning based' (2007: 275). Spada sees this as a misconception, and refers to the body of classroom-based research which has indicated the value of some measure of focus on form (see also Section 2.5.3.2.)

Parallel to these pedagogical changes, accounts of second language acquisition have been developed based on theories which see language learning taking place in the same way as other cognitive learning activities. At the same time those who favour the innatist accounts of L1 language learning continue to present their case (e.g. Hauser, Chomsky & Fitch (2002), who attempted to rephrase Chomsky's original hypotheses in a manner which tends towards finding common ground with those who previously held contrasting views). Attempts to find a universally applicable theory of language acquisition will undoubtedly remain the preoccupation of SLA researchers.

2.5.2   The Work of Pit Corder

As already noted, studies of SLA were, until the 1970s, built on the premise that all language learning could be traced to behaviourist systems in the learner where the influence of the mother tongue was paramount; in other words, second language learning was considered as the formation of habits and learners would characteristically carry over what they know in their mother tongue into their L2. Even before the beginning of that decade, however, the work of Corder in many ways presaged developments which became the focus of work done in the area of second language acquisition in the course of the following 30 years, the results of which provide the theoretical underpinnings for this thesis.

2.5.2.1 Input

The papers collected together by Corder in 1981 had all been previously published as separate articles between 1967 and 1980 (Corder 1981). Corder's remarks on the difference between 'input' (that which is made available during instruction, for example by a teacher) and 'intake' (that which the learner himself controls and takes in) were taken up by a number of researchers (e.g. Sharwood Smith 1993). Correlations, Corder suggested, between a learner's language and potential input would enable the formulation of descriptions of a learner's learning processes. An analysis of the LINCS corpus should contribute to this dimension of SLA.

While Corder's notion of 'successive states' of learner language (1981: 27), the original version of which appeared in 1971, was influenced by Brown's studies of L1 language development (e.g. Brown 1968; 1973), the cross-fertilisation of ideas at the end of the

1960s resulted in not only the seminal work on morpheme acquisition by Dulay & Burt (1974), but also in Krashen's Input Hypothesis (Krashen 1982). Corder also mentioned a 'built-in syllabus' (1981: 13) structure within the mind of a learner acquiring a language. Thereafter, eight years later, Pienemann's Processability Theory (PT) (Pienemann 1989) outlined the necessity to sequence instruction in a certain way in order to accommodate such a pre-existing acquisition order in the learner.

2.5.2.2 Interlanguage

Corder also notes that '(…) we come to the conclusion that the concept of ungrammaticality or deviance is not applicable to the learner. Everything he utters is by definition a grammatical utterance in his dialect' (Corder 1981: 32). A clear signal of what Selinker (1972) termed 'Interlanguage' (IL) can be seen in this statement, a term which has become an integral item in the SLA vernacular.

2.5.2.3 Formulaic Language

In L1 acquisition, Corder's work was preceded, as in other contexts mentioned in Section 2.5.2.1, by the work of Brown who, in the context of the acquisition of Wh questions, suggests that at the early stage children's questions are expressed as 'unanalyzed routines' (Brown 1968: 279).

One of the contentions presented in this thesis is that in the early stages of acquisition, some parts of the German passive are learnt as formulae or memorised chunks, as described by Myles for French (2004) and Wray for English (2007), for example. This notion of pre-learned, formulaic chunks, too, was presented by Corder in his discussion of the well-formedness of learner language and the importance of identifying errors correctly. 'Just because a sentence is acceptable is no proof that the learner knows the rules by which it is formed. He may have learnt it parrot-fashion as a whole or as a formula (...)' (Corder 1981: 41), he suggests.

2.5.2.4 Corpus Studies

Language is now often analysed by means of a learner corpus, and this, too, is hinted at by Corder. He indicates the value of studying learner language through this means in

the following way: 'The applied linguist's study of the learner's language is an attempt to characterize the "approximative system" of a learner (*or a set of learners*) from the data of his utterances.' (1981: 53, my italics).

Use of the term 'approximative system' is revealing in that it further substantiates the idea that a learner's language is not fully formed in terms of its proximity to NS norms. It is the intention in this study, carried out from the perspective of an applied linguist, to distinguish the characteristics of L2 German with respect to the syntactic means of expressing the passive, based on examples from the written production of a specific *set of learners*, or, in the present context, based on data collected in a learner corpus.

2.5.2.5 Error Analysis

It has been claimed that part of the learning process for a learner is to benefit from the feedback given by the teacher. In general terms, the accepted way of giving feedback is to indicate errors, whether or not they are also corrected. Error analysis was the main differentiating method of analysing language progression during the 1970s. Corder (1981) suggested that errors can only be identified through careful investigation of meaning which would be necessary in order to establish clearly the intent of the learner and thus the presence or absence of an error. Corder's distinction between *error* and *mistake* is adopted in this study, the former referring to 'systematic errors of the learner from which we are able to reconstruct his knowledge of the language to date' and the latter meaning an 'error of performance' and which occurs by chance as the result of a momentary lapse (Corder 1981: 10). The process of identifying errors in the corpus under examination here will be explained in Chapter 4 and will be followed by an investigation of the correctness of learners' production of the two syntactic forms of the German passive which are differentiated by their meaning.

2.5.3   Aspects of Instruction: Input and Output

This section reviews some accounts of SLA which are important to the evaluation of the learner language which is described in this study. Studies relate to several of the concepts mentioned in Section 2.5.2, indicating the relevance of the work of Corder (1981) to the framework in which this study is set. Firstly, the language input to a

learner is considered; secondly, the processes which may affect the output or production of the learner and thirdly the output itself.

The evidence outlined in Section 2.5.3.4 will indicate that an L2 learner's interlanguage exhibits specific features and that sequences of development are observable in second language acquisition. Accordingly, the working hypothesis proposed here is that progression-in-stages occurs in the production of complex constructions such as the passive. Written language produced by advanced learners of German as a second language provides the data to test this hypothesis.

2.5.3.1 Input

The role played by different types of input, the effectiveness of instruction and the contribution of metalinguistic knowledge in the acquisition of a second language provided the subject matter for a large number of studies.[15] Theories have continued to evolve and in some cases earlier contentions have been restated in the terms of more recent propositions, for example Sharwood Smith's work on input enhancement (Sharwood Smith 1993), based on the earlier term 'consciousness-raising' (Sharwood Smith 1981) and now reiterated under the Modular On-line Growth and Use of Language (MOGUL) framework (Sharwood Smith & Truscott 2005). While this framework considers the process of language acquisition from a theoretical stance and may or may not have implications for language pedagogy[16], VanPatten's Processing Instruction (PI) is a practical application of his own Input Processing theory, which encourages a learner to focus on a particular form, targeted explicitly in the classroom input, with the aim of promoting acquisition (VanPatten & Cadierno 1993; VanPatten & Williams 2007). The early iteration of PI contained the notion of the First Noun Principle, according to which learners interpret the first noun they encounter in a sentence as the semantic agent or grammatical subject. Evidence to support the validity of this notion has been advanced by Hikima (2010) in the case of the acquisition of the Japanese passive. It is possible to build on the study reported here and to test the principle in the future in a similar way via pre- and post-tests using a series of pictures illustrating active and passive events on a sample of L1 E learners of the German passive. VanPatten & Williams (2007: 116) also point to the fact that learners 'cannot

---

[15] The collection of papers in Studies in Second Language Acquisition (Lightbown, Spada & White 1993) provides a useful example. Two of them, Sharwood Smith (1993) and VanPatten & Cadierno (1993) are mentioned below.
[16] As yet no specific practical applications for this framework have been explored.

process and store the same amount of information as native speakers can during moment-by-moment processing'. This comparison of NS capacities with those of learners is useful for the study described here and will be mentioned again in the context of the data analysis in Chapter 6. Also of interest to this study in terms of pedagogical strategies for teaching grammatical items such as the passive is the fact that, for VanPatten and his associates a learner acquires grammar largely subconsciously, that is, implicitly, whereas for others an explicit focus on grammatical explanation is necessary for acquisition to take place (e.g. Long & Robinson 1998).

2.5.3.2 Form-focused Instruction

Long (1996) suggests that items of grammar should be explained in a classroom environment only when a query is raised by a learner concerning language used in that setting. In a slightly later version of the strategy, he and Robinson explain that '[*f*]*ocus on form* often consists of an occasional shift of attention to linguistic code features – by the teacher and/or one or more students - triggered by perceived problems with comprehension or production' (Long & Robinson 1998: 23). Grammar instruction should not be carried out in isolation from its context, which would fall into Long's category of Focus on FormS. In both cases, the primary aim in the language classroom is not focusing the learners' attention on particular grammatical features of the language but on conveying meaning.

Currently, the main tenets of FFI provide the basis for task-based learning (TBL), a teaching style which has evolved from the communicative method. In the TBL style, 'the language must come from the learners themselves' and the ultimate aim of a task is 'the communication of meaning within the classroom setting, not necessarily grammatical accuracy' (Cook 2008: 257).

It has been said that complexity of structure may have some bearing on the efficacy of teaching (Norris & Ortega 2001). Possibly more complex structures, such as the German passive, can only be taught, not least because of infrequent occurrence in the input. (Klapper & Rees cite Harley & Swain 1984, White 1991 and Lyster 1994 whose work is said to confirm this point.) Klapper & Rees (2003) underline the importance of the role of explicit instruction in a formal learning environment. This view finds support from Lightbown (1998) who writes of the positive effect of instruction carried

out at a point at which a learner is linguistically ready to acquire a grammatical function, in contrast to Krashen (1982) whose diametrically opposed view sees no such benefit.[17] Norris & Ortega's meta-analysis of 49 studies published between 1980 and 1998 confirms the apparently long-lasting effects of instruction, but the authors suggest that further proof is required because of a lack of rigour in research design (Norris & Ortega 2000: 500). While the studies previously mentioned in this paragraph deal with FFI from a teacher's perspective, from a theoretical standpoint Krashen (1982) sees focusing on form as only of use to a learner in monitoring production during, for example, a grammar test targeting a particular item or in a situation where time is available for reflection. By extension, in the study described in this thesis, it might be expected that use of the German passive may be variable depending on the learning situation. Because this learner corpus contains both time-constrained and free writing, it will be possible to look for examples of usage of the passive in both contexts and then to investigate whether there is any variation between them.

2.5.3.3 Developmental Sequences

As mentioned in Chapter 1.5, Myles has stressed the need for the collection of data which indicate changes in the learner's production, in order to investigate language development (Myles 2005: 374). With this in mind, the next section considers a number of studies which have treated the question of fixed-stage as opposed to continual development from a theoretical or empirical perspective.

If UG offers an explanation only of the representation of 'unconscious interlanguage knowledge' (White 2003: 37) in the human mind, then the developmental aspect of language acquisition as exemplified in production must be provided by supplementary theories as suggested below:

> To account for grammar change (i.e. development), one needs a theory of how the input interacts with the existing grammar, what properties of the input act as triggers for change, what properties force changes to the current representation, what might drive stages of acquisition (White 2003: 37).

---

[17] Note that the subject of the effectiveness of pedagogical intervention in the order of language acquisition is still very much open to debate.

Debates are still ongoing as to the extent to which a second language learner has elements of UG at his or her disposal, that is, the extent to which prior existing knowledge, whether it is innate (implicit) or learned (explicit) is involved in second language production. Differing points of view are based on the degree of access which a second language learner has to UG, e.g. the Full Transfer/Full Access model. This model presumes that UG in its entirety can be drawn on when learning a second language and that all the elements of the learner's L1 are carried over and are available for learning an L2, or to put it in generative terms, the end state of L1 acquisition is the initial state of L2 acquisition (Schwartz & Sprouse 1996). Whatever the theoretical standpoint, the fact that the L1 is already present when the learning of the second language commences suggests that even if the acquisition process is the same, resources are available to the L2 learner which are not present when that learner is acquiring the L1. The arguments in this study are not based on any single theory, although generative linguistics appear to offer the most credible explanation for L1 learners being able to construct language patterns of which they previously had no knowledge. Developments which attempt to combine aspects of more than one theory or framework will continue, in the ongoing search for a single unifying theory of second language acquisition.

What is generally accepted is that L2 development occurs gradually and is subject to some constraints even if it is not yet clear where these constraints originate. Among the questions this raises is how this developmental change should be classified. The idea of the structure of language changing incrementally over time has been the subject of several studies since the early years of SLA research. Brown's (1973) enumeration of the acquisition of grammatical morphemes in L1 is generally considered as the first reference point in the discussion of progression in stages.[18] Since that time, there have been a number of theoretical studies which have aimed to describe developmental patterns in learner language in terms of fixed stages or gradual continua.

For example, in the explanation of his Natural Order Hypothesis, Krashen enthusiastically welcomed the identification of developmental sequences in language acquisition in these terms:

---

[18] Note that there was no correlation between simplicity of form and ease, i.e. order, of acquisition: the third person singular morpheme –*s* is the penultimate step in Brown's (1973) acquisition order.

One of the most exciting discoveries in language acquisition research in recent years has been the finding that the acquisition of grammatical structures proceeds in a predictable order. Acquirers of a given language tend to acquire certain grammatical structures early, and others later (Krashen 1982: 12).

In the context of the present study, the variety of means of expressing the passive in German, which will be described in Chapter 3, may well demonstrate a hierarchical order of acquisition.

Other hypotheses advanced by Krashen, based on research into L2 acquisition both in instructed and in naturalistic environments, included the Input Hypothesis (Krashen 1982: 20), according to which the process of acquisition occurs incrementally, depending on a structure i already being acquired before the next stage ($i+1$) can be achieved. A similar progression, based on classroom research, is described by Pienemann's Teachability Hypothesis, suggesting that the next feature along the acquisition continuum can only be acquired once the previous one in the acquisition order is in place in the learner's grammar (Pienemann 1989). Pienemann subsequently proposed a five-stage sequence in the acquisition of grammatical structures, beginning with x (a simple lemma, or word, item), progressing through phrase and clause structure and ending with full sentence structure (Pienemann 1998). According to this Processability Theory (PT), a new construction may appear on a limited number of occasions in spontaneous production and is said to have emerged,[19] while other, pre-existing constructions remain dominant.

At first glance, the ranked order given by Brown for L1 (1973), the $i+1$ of Krashen (1982) and the $x + 1, 2, 3, 4, 5$ of Pienemann (1998) for L2 give an altogether mathematical impression of language acquisition suggestive of the completion of one stage before another begins. A similar mathematical expression illustrates the point in time when a new structure may be said to have been acquired: 'a rule is *not* part of the learner grammar at Time T and *has* been incorporated into it at time $T + 1$' (Sharwood Smith & Truscott 2005: 222). If such mathematical descriptions suggest a rigid, step-wise progression, it should be remembered that it is the steps which are clearly defined, and not the timing. Brown (1973), Krashen (1982), Pienemann (1998) and Sharwood Smith & Truscott (2005) all suggest that learners will acquire language at their own

---

[19] See Section 2.5.3.5.

individual pace; a more gradual process along a continuum exists which is characterised by the co-existence of NS-like and NNS-like forms (Sharwood Smith & Truscott 2005). This will be raised again in Section 2.5.3.5 and in Chapter 7.2.2. The pace and process of acquisition may be affected by a number of different factors. For example, even if learners receive input at a level just beyond their present one as Krashen suggests (i.e. *i*+1), it is not certain that acquisition will automatically follow because the learner may not be able to comprehend the input.[20] This may be due, according to Krashen, to the so-called 'Affective Filter' (Krashen 1982: 30) being raised so that uptake of input is prevented by, for example, an individual's reluctance to acknowledge input on a given occasion. A physical distraction such as classroom noise might also be responsible for learners not taking up the input and therefore failing to make the progress for which instructors aim.

In his study of the acquisition of German word order by three eight-year-old Italian girls, a precursor to PT (Pienemann 1998) which could be applied to all languages, Pienemann (1980) refers to the overlapping of acquisitional stages; in other words, there is no clear defining line between one stage in word order development and the next. As has been noted, a learner is said not to be ready to acquire the next stage until prerequisites from the previous one have been assimilated. The idea also has resonance in construction grammar theory, which suggests that it may be easier for a learner to acquire a new construction if that construction has already been partially 'supported' (Abbot-Smith & Behrens 2006: 995) by prior evidence in the input and has led to acquisition. Abbot-Smith & Behrens found that a child acquired the *sein* passive before the *werden* passive because he had previously acquired other, similar constructions (such as *sein* used as a copula verb).

Pienemann's PT (Pienemann 1998) proposes both a developmental and a variational aspect of language acquisition: it evolved from the Multi-dimensional Model developed during research on the *Zweitspracherwerb italienischer, spanischer und portugiesischer Arbeiter* (ZISA) (= The acquisition of a second language by Italian, Spanish and Portuguese workers) project in the 1970s. Certain elements in a learner's language will develop according to fixed patterns, while others which have no such fixed order of

---

[20] Note Krashen's definition of 'acquired' in contrast to 'learned': the first is subconscious and implicit, the second conscious and explicit. There is no interface between the two systems, which cannot be interchanged (clarified by Ellis 1994).

acquisition will appear in a random fashion and may respond positively to instruction. The developmental stages outlined by Pienemann cannot be changed, though he admits that pedagogical intervention can at least increase regular use of a structure if a learner is exposed in practice to items which are present in the language of the next developmental stage. Krashen (1982), on the other hand, maintains that only comprehensible input, whether provided in a formal teaching situation or elsewhere, is converted into acquired language; one of the criticisms levelled against this theory (the Input Hypothesis) is that it does not specify exactly what counts as comprehensible. The small C corpus which forms part of the data collection described in this study is included as an example of the kind of input which would be comprehensible to the learners in the A cohort. Most of the texts within it were chosen by the tutor as reading material in preparation for exercises in written production, judged by her to be pitched at a suitable level. In Krashen's terms, this level would have to be slightly ahead of the learner's current competence in order for acquisition to take place.

2.5.3.4 Interlanguage

Interlanguage, by its very nature, can be different but no less correct in the learner's mind, at every stage of its use and even on a daily basis. Although the definition of the concept is generally attributed to Selinker (1972), it has already been pointed out in this chapter (Section 2.5.2.2) that it had previously been used to describe learner language by Corder in 1971. Selinker's elaboration of the term not only suggested that IL is a hybrid between a learner's L1 and the target language (TL), but that, in coming between two languages, it has its own identity, with a system of patterns which may not be found in either the L1 or the L2. It is essentially an interim step towards the supposed full, final acquisition of a second (or subsequent) language and is thus in itself a stage in language development. In a later paper, Lakshmanan & Selinker describe IL as 'a linguistic system in its own right' (Lakshmanan & Selinker 2001: 395). Corder, too, suggested that the language a learner produces is at all times part of a system, albeit that system does not conform fully to the norms of the TL (Corder 1971). The interlanguage system evolves because of a dual process: the learner both receives language (in the input) and produces his own systematised version in production. Some of this production will be consciously monitored by the learner, depending on whether he is understood or corrected. 'One must allow for the possibility that the learner may utilize at least two interlanguage grammars, one for productive use and one for receptive use'

(Corder 1981: 75). More recently, describing IL as a developing system, a system in transition, Pallotti (2007: 362) said, '[T]oday's interlanguage is qualitatively different from (…) that of yesterday'. This study will seek to identify transitory phases in the acquisition of the German passive, showing L1 E learners' attempts to move from 'novice' to 'expert' in Ellis's terms (Ellis 2005: 333), although no firm claims regarding the sequence of acquisition can be made based on a corpus of a small size and consisting of a variety of text types.

The danger of regarding IL as somehow inferior to the target language assumed to be achieved by all native speakers was also pointed out by Bley-Vroman. According to his Fundamental Difference Hypothesis (FDH), learner language should not be considered 'a degenerate form of the target system' (Bley-Vroman 1983: 4). The temptation to compare NS language with learner language, using the former as a benchmark, leads to the comparative fallacy as identified by Bley-Vroman (1983), and should be avoided. It should be remembered that this type of comparison is difficult, if not impossible, for most teachers to avoid, since study assessments are to a large extent centred on a learner producing non-NS-like language. Language teachers should take their inspiration from SLA researchers in the field of applied linguistics in order to determine the best way of implementing the findings of the SLA research in a pedagogical context.

Lakshmanan & Selinker (2001: 400) used the term 'interlanguage output' to refer to spoken data samples, but in this study it will be applied to written data only, and will assume that the avoidance of the comparative fallacy applies also to this type of IL output. Lakshmanan & Selinker (2001) argue further that observations regarding the current state of interlanguage should be made not only with reference to a particular IL output but also to an investigation of NS output on the same task. They suggest that ungrammaticality, as judged against NS norms, may even be evident in NS output so that measuring IL only against the TL could result in an invalid judgement. By comparing the written output of one NS of German with that of several learners with L1 E on the same tasks in this study, it should be possible to discover whether there is evidence to support this sugggestion. In a more recent study, Dekydtspotter, Schwartz & Sprouse (2006) argue against the 'strongest form' of the FDH (2006: 33), suggesting that L1 learning and L2 learning are subject to the same processing systems and not to different ones, as proposed by the FDH.

Cook (1996: 6) uses the term 'independent language assumption' for IL, which embodies the ideas of the individual, his language and his reason for producing it. The aim of communicative language teaching to which we referred earlier is, as the name suggests, to allow the learner to communicate with other speakers of the TL: the TL norms do not have to be respected in order to achieve this and therefore the IL should not be judged by those norms, even though learners and teachers will be hoping eventually for as close an approximation to the TL as is realistically possible in the context of a formal learning environment.

Bialystok & Sharwood Smith's (1985) notion of IL being composed of two separate elements, 'knowledge' of language and the learner's subconscious ability to analyse it and 'control' of language and the learner's capacity to blend required elements in production, reflect the generative (Chomskyan) perception of 'competence' (what a learner knows about language) and 'performance' (how a learner uses that knowledge to produce language) (explained in Cook 1993). In Bialystok & Sharwood Smith's view, both of these elements progress during IL development. The focus of their article is, like the majority of IL studies, spoken language which by virtue of its spontaneity is usually considered the most suitable source of language acquisition data, being less constrained, if not entirely unconstrained by conscious influences. Spontaneous speech should have communication as its primary objective and may include several instances of non-native grammatical usage, according to Pienemann (1989). It is therefore more likely to represent the current level of the learner's intuitive grammar (his or her 'competence' in Chomskyan terms) than is language which has been produced as a result of conscious thought processes or with the support of metalinguistic knowledge (his or her 'performance'). It might seem, then, that only spoken data would provide reliable evidence from which to draw conclusions concerning the current state of a learner's IL. On the other hand, many teachers would testify to having heard a speaker apparently being constrained by some kind of Monitor (Krashen 1982) on detecting perceived imperfections in his or her spoken performance and re-phrasing incorrect utterances resulting in what can be termed 'constrained speech'.[21] In such a case, evidence from spoken data might be thought to be inconclusive. Norris & Ortega (2012: 582) discuss the relative validity of naturally-occurring (i.e. spoken) data as

---

[21] Krashen's Monitor Hypothesis posits that monitoring language use depends on a learner having a number of things at his or her disposal: time, the will to focus on form and knowledge of the rule to be applied (Krashen 1982: 16).

opposed to production data elicited in written exercises and conclude that while spoken data may be regarded as representing more naturalistic language use, an advantage of collecting written data lies in the possibility of controlling some of the variables which are inherent in spoken exercises such as external noise factors and nebulous learner interactional features. That is not to say that more variables can be controlled in one type of assessment than in another – what is of key importance is that the assessment fits the aim of the study. Myles (2005) points out that written data do have their place in acquisition research in particular when it is intended that the results should be applied to language teaching, and a learner corpus of written German would therefore seem to offer a suitable data sample in the investigation of the relationship between SLA and second language pedagogy, in the terms of research question 1 set out in Chapter 1.7.

The next four sub-sections contain observations concerning terms which are useful in characterising learner language and which may be helpful in assessing the interlanguage which is being examined here, firstly acquisition and emergence and thereafter formulaic language, optionality and, lastly, fossilization.

2.5.3.5 Acquisition and Emergence

With reference to the variety of ways of defining the point at which something is acquired, it is noted that Mills (1985) uses 'appear', suggestive of first occurrence, whereas conversely Eisenbeiß (1993: 241) speaks of 'frequent and productive use'. Abbot-Smith (2003: 171) concludes that it is not possible to 'draw a line' and consider an item acquired. Most researchers agree that after a single correct occurrence of a structure, there will be occasions on which an incorrect version is produced, and if we follow Sharwood Smith & Truscott, for example, the specification of a point of acquisition is unnecessary (not to say more or less impossible): language development according to the MOGUL model, mentioned in Section 2.5.3.1, will occur both incrementally in steps and simultaneously along a continuum until, in MOGUL terms, the might of the correct structure prevails over its weaker, incorrect competitors, resulting in general usage (Sharwood Smith & Truscott 2005).

An important distinction is made between acquisition and emergence. As Gass (2013) points out, a number of different criteria have been used to define the point of acquisition. Krashen, for example, required a structure to be used 90% of the time

before its acquisition is confirmed (Krashen 1982), whereas Vainikka & Young-Scholten's benchmark was 60% (Vainikka & Young-Scholten 1994). On the other hand, Gass notes that 'one should not lose track of the important […] factor of emergence' (2013: 73) in order to measure developmental steps. Pienemann (1998: 138) sought to clarify the point of emergence by defining the Emergence Criterion. A grammatical element is considered to have emerged when it has appeared in merely a few obligatory contexts in the learner's output. Pallotti (2007) attempts to define a working model of Pienemann's Emergence Criterion and concludes that the problem of determining emergence centres on whether a learner is considered to be exhibiting systematic or random use of a structure.

Research studies which seek to establish whether or when an item has been acquired often measure the number of times the item is used correctly in an obligatory context, calculated on a percentage basis. In Chapter 3, where the various means of expressing the passive in German are discussed, it will be evident that it is not always possible to determine what constitutes an obligatory context in the use of this construction because there are options available to users of German which enable an active sentence or a reflexive verb to replace passive morphology. In the study described here, an alternative measure of accuracy will be introduced in Chapter 4, but it is useful at this stage to point out the difficulties inherent in attempting to establish whether or not a particular item has been acquired.

In this small sample, the term 'emerged' is applied to an item whose grammatical form has appeared 5 times in the learner's language, to take account of a number of guesses or formulaic use, while a benchmark of 80% accuracy in aligning grammatical form with semantic function is proposed as the requirement for considering an item acquired, for reasons that are elaborated in Chapter 6.3.

2.5.3.6 Formulaic Language

A body of research in the functional perspective into the acquisition of early stage language which appears to be based on a learner's use of formulaic expressions is growing exponentially as the volume of language stored in learner corpora becomes widely available. Wray says that learners work on the basis of a 'needs-only analysis' (NOA) (Wray 2007: 873) and that 'class-room learners rely heavily on memorised

wordstrings' (Wray 2007: 875). If *sein* passives are expected to occur so infrequently in German texts as Duden (1998) indicates, then it is possible that learners may not devote much time to ensuring that they are produced in a NS-like way, preferring to opt, using an NOA, for the more common *werden* passive as a coverall, especially as the *werden* passive is the one taught as 'the passive' to learners with L1 E as in the example teaching method described in Section 3.6.

Myles (2004:145) identifies grammatical structures such as simple means of expressing personal preference produced by beginner learners of L2 French as typical of the wordstrings mentioned by Wray (2007: 875). These structures are not evident in any of the language produced outside the chunks, indicating that these learners have memorised the chunks without being able to further manipulate the grammatical structures embedded in them. L1 E learners of German may memorise the most straightforward type of *werden* passive, a 3pers form + regular PP as explained in Chapter 3.3.1.

Wray (2008) points out that formulaic language remains difficult to define and suggests that individual researchers may refine her (Wray's) definition depending on their own research aims. In addition, she remarks that her descriptions of formulaic language use sit less plausibly with generative studies than they do with other research agendas, such as usage-based studies, for example. This may be particularly relevant in studies of the acquisition of aspects of syntax such as the one described here. A generative approach attributes grammatical development to pre-set, innate properties and separates syntax from semantics, in contrast to a functional, phraseological approach in which language is acquired following many hours of exposure to meaningful word units (Ellis 2008). As has already been stated in Chapter 2.1, this study draws on theories of acquisition from more than one theoretical stance, and for present purposes, it is proposed that the definition provided by Wray, although more usually applicable to her theoretical approach, offers a suitable means of categorising some of the learners' output which is examined here.

Wray defines a formulaic sequence as 'a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar' (Wray 2002: 9). It remains her benchmark in

subsequent work e.g. Wray 2008; 2013. The application of this definition to the study described here is considered during the evaluation of the quantitative and qualitative data in Chapters 5, 6 and 7.

2.5.3.7 Optionality

In Section 2.5.2.5, the need to investigate the intent of the learner before ascribing an error to his or her production was pointed out.[22] Identification of intended meaning has become a clear focus of SLA research, for example Sorace's account of optionality (Sorace 2000), which predicts that two or more forms of the same semantic item, using the same lexical resources, may still be present in the language produced by very advanced learners. In the case of the German passive, learners may continue to use the *werden* and *sein* passives interchangeably to equate to the English 'be' and because they remain unsure of the semantic difference in German between the two passive auxiliaries. Papp (2000: 195) suggests that optional forms remain a possibility in non-native grammars particularly if the constructions to be learned are scarce or rare.[23] She remarks that '[w]here data about a certain parameter of grammar continues [*sic*] to be very scarce, the marginal state will persist into adulthood or at the otherwise near-native level of attainment'.

This may well be the case for learners of the German passive, given that it occurs comparatively rarely (only 7% of all finite verbs in contemporary German texts, according to Duden 1998). The notions of divergence, convergence and incompleteness used by Papp (2000: 177) to differentiate various levels of competence will be useful terms of reference when evaluating the learners' written production in this study in Chapter 8. They suggest that a learner has NS-like expressions as his or her aim in the output and that these are sometimes but not always present. If the study of the interlanguage of the A cohort of students, reported in Chapter 7, reveals the co-existence of such forms, it will support the claim that 'a new solution does not force the extinction of an old one in a straightforward [i.e. rigid as opposed to gradual; my clarification] manner' (Sharwood Smith & Truscott 2005: 221).

---

[22] The notion of 'passive intent' is introduced in Chapter 4.
[23] Optionality is also observable in L1 language use.

Papp refers in her 2000 article to an area of study which has since become centred on the Interface Hypothesis (IH). This hypothesis has been described by Sorace & Filiaci (2006: 340) and elaborated in these terms:

> [N]arrow syntactic properties are completely acquirable in a second language, even though they may exhibit significant developmental delays, whereas interface properties involving syntax and another cognitive domain may not be fully acquirable.

Applying this hypothesis to the study described here, it may be expected that acquisition of the narrow, basic syntax of the German passive is relatively simple for a learner, but that in some cases production may not ever reach the optimal (NS) state because of the conflicts of meaning at the interface between syntax and semantics or syntax and discourse/pragmatics (Rankin 2011). Learner preference for a particular syntactic element may persist once the basic grammar (e.g. in this case the morphology of the auxiliary) has been acquired, because of L1 influence (e.g. in this case the *sein* passive is preferred by some because the English passive is based on 'be'). That is to say that some learners may never achieve full mastery of the production of the passive even if they know how to form it. If the German passive is taught to L1 E learners as a combination of an appropriate tense of *werden* with PP (see Chapter 3.2), it is likely that the two 3pers forms *wird* and *werden* will be the first options when a passive construction is intended, rather than a form of *sein*.[24] Difficulties in differentiating between the actional *werden* and the statal *sein* (see Chapters 1.3 and 2.2.2.2) may be due to a learner's inability to reconcile German passive syntax with German passive semantics. The results of the data analysis should shed some light on this claim.

2.5.3.8 Fossilization

Corder claimed that interlanguage development comes to a halt once a learner has sufficient language at his disposal to be able to communicate adequately. He described '(…) the phenomenon of "fossilization", where a learner's interlanguage ceases to develop however long he remains exposed to authentic data in the target language'

---

[24] The 1pers and 2pers forms (in other words, personal references), are in any case rare and therefore less likely to be found in academic writing. The 3pers forms are taught first in pedagogical textbooks (see Chapter 3.6).

(1981: 74). This somewhat broad generalisation is echoed in Selinker's Fossilization Hypothesis (Selinker 1972) and was subsequently revisited by Han & Odlin (2006) who differentiated between global and local fossilization. The latter applies to the stagnation of individual grammatical items, whereas the former applies to the interlanguage as a whole.

In the most recent review of fossilization, carried out 40 years after the term was first defined by Selinker, Han describes it as 'an interlanguage-unique phenomenon in which a semi-developed linguistic form or construction shows permanent resistance to environmental influence and thus fails to progress towards the target' (Han 2013: 133). It remains to be seen, following longitudinal analysis of the data, whether or not the term can be applied to the acquisition of a construction such as the passive which occurs only rarely in the naturalistic input.

As recently as 2013, Han has pointed to both L1 influence and L2 input as 'the core drivers of (adult) L2 acquisitional outcome' (Han 2013: 145). To the extent that Han believes that fossilization can only be examined in spoken, natural language and not in an FL classroom, her views may be of limited application to the present study, but some of her predictions nevertheless mirror suggestions that have already been made here, in the section on optionality, for example. Her first prediction runs as follows:

> [C]ertain TL usages are by default susceptible to fossilization because the input received by the learner is never going to be strong, because of their rarity and the intrinsically variable relationship between form, meaning and function (Han 2013: 145-146).

At this stage, it is worth noting that in some circles the term 'fossilization' is now rather unfashionable and 'entrenched errors' is favoured (Roberts 2013); this certainly appears to be less restrictive than the permanence which is inherent in Han's definition.

2.5.3.9 Transfer

The notion of transfer, that is, transfer of language systems from a learner's L1 to his or her L2, first received attention during the 1960s when it was thought that structures which were present in L1 would initially be carried over into L2. If a structure was

subsequently found by the learner either not to be the same or not to exist, then a process of conscious re-evaluation would have to occur before the learner produced the NS-like variation in the L2.

Cross-linguistic influence (CLI), the term which succeeded 'transfer' after it became unpopular in the 1970s because of its association with behaviourism, has been the subject of debate over several decades since Lado (1964) suggested that deviance from the L2 'norm' was caused by interference from L1 (see Section 2.6.1). On the one hand transfer was thought to be more likely where languages are typologically similar, such as English and German (see the typological comparisons in Hawkins 1968), so that more *sein* passives (related to 'be') than *werden* passives might be found in L2 German adult language. This would, however, sometimes result in negative transfer and the production of a non-native-like form.[25] On the other hand, more recently, Pfenninger, for example, maintains that acquisition of an item depends on the learner's ability to process that item, and that therefore 'productive knowledge and grammaticality judgements cannot be explained in terms of the structure of their L1' (Pfenninger 2013: 143).

In Chapter 3 of this study, the links between form, meaning and function in German passive syntax will be identified and compared to the same construction in English. At this juncture, it should be noted that the construction is likely to be less easy for NSs of English to acquire if they are attempting to transfer the system from their L1 for two reasons. Firstly, German has more ways of expressing the passive than English, so learners of German as L2 with English L1 may be confused because they cannot easily identify an equivalent construction in the target language. Secondly, and more importantly for the study described here, one of the syntactic forms requires the German equivalent of 'be' but is neither the most common form of the construction nor the equivalent in meaning of a 'be' passive, as explained in Chapter 1.3. If Corder's suggestion that interference from the L1 is more evident in adult L2 learners than in children (1981: 74), then it might be expected that adults with English L1 learning German as L2 would opt for the auxiliary *sein* (= be) rather than *werden* (= become) in order to express the passive. It could be argued that learners may take the easy way out

---

[25] Negative transfer inhibits the acquisition of a linguistic item, whether phonological, semantic or syntactic, whereas positive transfer facilitates acquisition (Isurin 2005).

and seek a solution to their production problem by resorting to their L1 rather than resolving a tricky quandary (Sharwood Smith 1996).

Following the data analysis, the part which cross-linguistic influence may have to play in the acquisition of the German passive by learners with English L1 will be raised again in Chapter 8. The next main section of this present chapter is concerned with the ways in which the type of language that has been described in sections 2.2 and 2.3 has been analysed, indicating the transition from the examination of two language samples in contrastive analysis to the opportunities afforded by larger corpus samples and, with particular relevance to the study described in this thesis, by learner corpora.

2.6   Contrastive Analysis, Contrastive Interlanguage Analysis and Learner Corpora

While selected elements of linguistic and SLA theory and their application to this study have been explored in sections 2.2 and 2.5 of this chapter, it remains to examine the literature on the methodological foundations on which the treatment of the collected data is based. Learner corpus research is a comparatively new discipline, having its beginnings only in the 1990s. Literature on the subject is therefore less abundant than that in other research fields with greater longevity. This sixth section of Chapter 2 indicates how learner corpora have developed from contrastive analysis techniques in various disciplines since the 1960s and refers to examples of existing learner corpora in foreign languages and in English which were pioneers in their field, to indicate the background against which the LINCS corpus has been collected.

2.6.1   Contrastive Analysis

Contrastive Analysis (CA) developed over a decade from what had been an avenue of enquiry based on what Granger (1994: 37) termed a 'slim empirical foundation' into a robust and universally acknowledged methodology, not only in the field of translation studies where parallel corpora containing original and translated texts enabled bi-directional study (e.g. Granger-Legrand 1976), but also for example in that of FL pedagogy. In the 1960s, Lado's work on contrastive linguistics provided the rationale for predicting the difficulties which learners might encounter, depending on the similarities between their L1 and the L2 in question. He expressed the following view:

Of special interest to the language teacher is contrastive linguistics, which compares the structures of two languages to determine the points where they differ. These differences are the chief source of difficulty in learning a language (Lado 1964:15).

According to this view, learners whose mother tongue is English, where the passive is constructed for the most part with but a single auxiliary, and who are learning the German periphrastic passive with its two auxiliary forms, should indeed find the construction difficult to acquire.

In FL pedagogy, the main focus of CA during the 1960s was predictive; in other words, a hypothesis could be produced to indicate or predict what might be problematic for learners of language Y (e.g. the inclusion of subject pronouns in English) with native language X (e.g. Spanish, where subject pronouns are not obligatory) and could be tested by comparing samples of linguistic items in languages X and Y. The behaviourist view of language learning which prevailed at that time was founded on the belief that L2 learners would transfer patterns from their L1 into their L2. In terms of the Contrastive Analysis Hypothesis (CAH), the habits formed while learning the L1 would dominate in the initial learning of the L2 and would show up as interference. Thus, if an L1 structure differed from an equivalent structure in the L2, the learner would initially use that structure in the same way as in his or her L1 until experience with the L2 structure overcame the weaker L1 version. Accordingly, in the context of the present study, learners of German L2 with English L1 might be expected to use *sein* to express the passive because it is the equivalent of 'be'. It was on the notion of transfer from L1 to L2 being the source of learner deviance from the NS norm that Lado (1957; 1964) focused.

The importance of the transfer of systems of the L1 in L2 learning subsequently fell out of favour when it became apparent that errors could be attributed to rather more than transferring an item from a learner's L1 (e.g. Dušková 1969),[26] but the issue is once again receiving attention as mentioned in the section on fossilization.

---

[26] Dušková (1969) found in a study of 50 Czech learners of English that errors were more likely to occur if a particular grammatical structure did *not* appear in the L1.

The examination of learners' errors became the focus of empirical study during the 1970s. This decade saw the introduction of the concept of interlanguage, as explained in section 2.5.2.2. By looking first at the learner errors in the collected data, the specific areas of difficulty for individual learners could be identified and the problem diagnosed. Whether or not the identification of errors in a learner's language production is useful depends on the aim of the study in which that technique is applied. Error identification is not the primary focus here, where the significance of the presence or absence of the passive in the learners' interlanguage is explored.

2.6.2   Contrastive Interlanguage Analysis

After the end of the 1970s, and following two decades when corpus linguistics as a methodology was less widely used, it regained popularity as the Chomskyan view that corpus evidence alone was insufficient, indeed inappropriate, for the study of language acquisition failed to find agreement among those researchers who favoured usage-based theories. In addition, as McEnery & Wilson suggest (1996: 17), regularised use of the computer in many aspects of academic as well as personal life had a marked effect on the storage of data. The existence of electronic corpora signalled the beginnings of additional corpus types, primarily as a means of comparing an original text with a translated version of the same text, or of comparing two or more original parallel texts. The publication of the first Collins COBUILD English dictionary in 1987, a collaboration between the University of Birmingham and John Sinclair (whose views on the limitations of corpus tagging are mentioned in Chapter 4.4.3), exemplified the innovative use of corpora in new ventures.

A new model of contrastive analysis, Contrastive Interlanguage Analysis (CIA), which developed from CA, was introduced to the field in 1994. Instead of bilingual translated corpora providing the data for analysis, examples of learners' interlanguage are collected with the intention of providing a richer supply of evidence in the analysis of language development in order to 'discover the features of nativeness and non-nativeness of learner language by comparing it with native language' (Gilquin 2001: 96). This accords well with this study's research question 4 which aims to identify similarities or differences in NS and NNS language. CIA allows studies to focus on comparisons not just between 'two different languages but between native and learner

varieties of one and the same language' (Granger 1996: 43).[27] This latter combination provides the data for the A and B corpora in the study described here.

Whereas CIA is more usually used to investigate the degree of L1 transfer from one language to another (Gilquin 2001), it would seem that a close comparison of the interlanguage produced by a number of different learners of the same language with the same mother tongue, as in the LINCS L1 E cohort, would indicate the stages through which adult learners of the German passive may be hypothesised to pass. In the present study, it is suggested that the development of the German passive in the interlanguage of learners with L1 E follows a path which is both similar to all but subject to variation between individuals. An examination of NS usage of the German passive is included because, from a practical point of view, in an L1 E pedagogical setting, correctness, benchmarked against a NS norm, is the ultimate aim for learners. The learner language is considered here in a positive light as systematic usage and not as an inferior variant of the TL. In addressing research question 4, any observed differences between the production of the passive by NSs of German and that which is exhibited by learners of L2 G with L1 E will be reported, and ultimately reasons for the differences will be suggested.

A CIA approach will therefore be adopted in the present study where the IL produced by a number of L2 learners of German with English L1 is being compared, and where a sample of NS German is available as a control in the B corpus. All except one of the subjects in the A cohort of this project have L1 E, while the B corpus comprises the written language of students with L1 G. CIA can therefore be implemented in a comparison of the two. In addition, a variation on CIA can be carried out by studying the interlanguage of the same cohort of learners both diachronically in a longitudinal analysis and synchronically in a cross-sectional analysis.

Granger (1996) and Gilquin (2001) both emphasise the predictive power of CIA studies based on the process of collecting and subsequently analysing bodies of text before arriving at an hypothesis because of the amount of data which can be examined, thus

---

[27] See Lozano & Mendikoetxea (2013) for a recent example of a corpus which is designed to facilitate CIA in a similar way to the present study. A possible limitation of the method is outlined by Gilquin (2008) who points to the need for equivalence in the text type of the NS versus NNS evidence. In the present study, the aim is to observe similarities or differences in the written language of approximately similar academic cohorts exhibited in two corpora of similar sizes (see Chapter 7.1).

making it a more reliable avenue of research with a broader empirical base than CA. However, CIA can also be used to test hypotheses as well as to generate them. The present study uses the former strategy: predictions will be derived from SLA theories such as those described in Section 2.5 and the collected data will provide evidence for or against the hypotheses generated rather than acting as the base from which they are constructed, although further questions may well evolve once the database has been consulted.

The machine-readable data collections to which this type of analysis can be applied are now described in Section 2.6.3. Examples of existing learner corpora with which the LINCS corpus may be compared are presented.

2.6.3   Learner Corpora

Once large datasets collected electronically were readily to hand, new dimensions could be added to fields of study: contrastive studies could be extended from translated parallel texts to examining similarities and differences between texts of other types. With the availability of large empirical samples of language came the possibility of examining not only the interlanguage of learners of the same second language with the same mother tongue, but also the interlanguage of learners of the same language but with different mother tongues. In the study described here, samples from both NSs and NNSs of German are collected in order to compare them, as explained in Section 2.6.2. It should be noted that the early claims were all based on English as the learner language. It is still the case that the greatest number of learner corpora focuses on English as a second or foreign language, as will be seen in section 2.6.3.2. For this reason the present study offers a contribution to the discipline, by providing an additional corpus in German.

2.6.3.1 Definition of Learner Corpora

The definition of learner corpora as 'electronic collections of spoken or written texts produced by foreign or second language learners in a variety of language settings' is taken from Granger (2002: 124). In the case of the present study, the written texts of advanced learners at university level in a formal i.e. not naturalistic environment, are collected in a number of settings, from short and longer free-writing exercises to

examination answers produced under time constraints. Pravec (2002: 81) suggests that '[g]enerally, learner corpora are important because they provide a deviation from the standard i.e. the language of the native speakers of a particular language'. On the basis that there is a contrast between learner language and NS language, this contention serves to underline the relevance of a learner corpus approach in this study.

Support for electronic corpus collection and analysis is provided by Adolphs (2006) who described multiple spheres of application such as language variation and comparing and analysing language varieties. Adolphs further remarked that '[c]orpora (…) are designed to represent a particular language variety' and points out that researchers must be particularly attentive in their choice of data to ensure close comparability between the research question and the empirical evidence on which it is based (Adolphs 2006: 17). With its focus on language development and its dataset comprising written language from three different sources at various levels of study, or in Adolphs' terms, language of three varieties, it would seem that the choice of a learner corpus as a database is justified in the attempt to answer the research questions which were identified in Chapter 1.7 of this thesis.

2.6.3.2 Existing Learner Corpora

Since their inception, learner corpora have traditionally been used to examine the performance of learners of L2 English, and to compare it to language from native speakers of other languages, first of all to indicate to learners specifically which patterns to imitate and which to avoid and secondly to indicate particular language patterns both to language instructors and researchers.[28]

Despite being a growing field of enquiry over three decades, Gilquin points out that the first volume of learner corpus studies, an edited collection of articles, did not appear until 1998. Four years after that date, Pravec's review article listed only 10 corpora then 'currently existing' (2002: 82). These were categorised as either Commercial (two) or Academic (eight) and the learners were mainly learners of English as a second or foreign language. In terms of relevance to the study described here, it is noteworthy

---

[28] Corpus studies more generally have served a similar purpose and indicate the value for teaching purposes of a large volume of data collected in a variety of situations e.g. the use of expressions of gratitude as formulaic sequences as described by Schauer & Adolphs (2006).

that the list on the Université Catholique de Louvain (UCL) website of corpora held by institutions or individuals other than the International Corpus of Learner English (ICLE) currently numbers 124, covering 12 languages. It is evident from these figures that the number of collections has increased considerably over the last decade. The same can be said of the variety of tools with which to analyse them. Some of these tools are identified in Chapter 4.

Because the system of collecting data in learner corpora is relatively young and their original aim had a different focus, they have not commonly been used in second language research; the present study seeks to expand the discipline in this respect. In the next section, four examples of learner corpora are mentioned for the purpose of comparison with the study described here.[29] A description of the design and aims of the Child Language Data Exchange System (CHILDES) is followed by comments on the French Learner Language Oral Corpora (FLLOC) and their Spanish counterpart, the Spanish Learner Language Oral Corpora (SPLLOC). An overview of the ICLE concludes the section.

The express aim of CHILDES is to gather together data and to make them available for a research public (MacWhinney 2000). It is mentioned here because it was designed specifically for research purposes and now offers the opportunity to SLA researchers to apply its analytical systems to their own corpora and, in return, to add their data to the public domain. CHILDES would therefore be considered a suitable repository for the LINCS corpus in due course. Held in electronic format in Carnegie Mellon University, Pittsburgh, USA, access to it is free and users are encouraged to contribute their own corpora to the collection, subject to a number of conditions of use. It has expanded since 1984 from a collection of spoken child L1 language in English and now contains a large volume of both spoken and written data in many languages at several levels of competence.

The number of corpora in German which have been contributed to the CHILDES system has doubled in the last few years but still stands at only 12. These corpora contain evidence from individual and multiple subjects, the majority of whom are below the age of four. All of them contain transcripts and/or audio recordings of spoken child language rather than L2s, and were collected primarily for the purpose of documenting

---

[29] A preliminary account of three of these corpora appears in Thoday (2008).

L1 development and to assist research in socio-cultural and socio-linguistic contexts through the elicitation and subsequent description of spontaneous speech (e.g. Szagun 2002). None of the German corpora were originally intended to be used for pedagogical purposes. Thus, they differ in content, in age of participants and in original purpose from the corpus of L2 learner language which is the focus of the study described here. Although Myles endorses use of the CLAN (Computerized Language Analysis) and CHAT (Codes for the Human Analysis of Transcripts) tools as being relatively easy to adapt for SLA research (Myles 2005), for the reasons advanced in Chapter 4, annotation was not deemed necessary for the current study. The possibility of transcribing and coding the LINCS corpus for inclusion in the CHILDES database remains an option for the future, and it is acknowledged that the benefits of shared data analysis systems are self-evident in terms of replicability and retrievability.

The French Learner Language Oral Corpora, which were the first foreign language learner corpora to be set up at the University of Southampton in 2001, made use of the CHAT tool to transcribe and code speech utterances for analysis. The field expanded relatively slowly: the Spanish version was not developed until 10 years after its French counterpart. In contrast to the present study, the data for these two corpora were collected from learners at secondary schools and the first studies contributed to assessments of learner performance and language development at this level. Subsequently, FLLOC and SPLLOC have been expanded to include data from university students. The present study is intended to shed light on the same educational sector, but it focuses on the acquisition of an aspect of syntax through the examination of written data. The teams who devised these oral corpora have not developed a similar corpus in German.

The ICLE, held at the UCL in Belgium serves as a prime example of a large dataset of learner language both in its form and in its intention. It is mentioned here because of its comprehensive contribution to the field. Containing the written essays of higher intermediate to advanced learners, it is similar in level to the corpus which has been collected for the present study, although the language of study is different. It was designed to 'uncover factors of "foreign-soundingness" in learner writing' (Granger 1996: 43). The addition of a list of corpora from around the world, together with further innovations, has made the UCL website a useful point of reference for researchers (Université Catholique de Louvain Centre for English Corpus Linguistics 2013).

Further observations on this list appear in Chapter 4.3 where the relative lack of corpora in German is assessed.

2.7   Review of Chapter 2

Chapter 2 has offered an overview of some aspects of the three main components of this cross-disciplinary study: theoretical linguistic descriptions of the passive, relevant theories of SLA with respect to language teaching and a description of the historical aspects of learner corpus research.

This overview provides the theoretical basis on which the study described here is founded. In particular, the data provided by the corpus in the present study (described in Chapter 4.6) fit the definition of a learner corpus given in Section 2.6.3.1 (Granger 2002: 24). It is necessary to reiterate, however, that no concrete conclusions can be drawn from a limited number of examples. Nevertheless, it is expected that learner corpus research can be furthered by this study which examines the language produced by learners in a variety of written exercises at a number of levels, albeit that the sample is not extensive. A study mentioned in Section 2.5.3.1 (Hikima 2010), for example, reviewed only 10 subjects.

The final section of the chapter contained an assessment of the contribution which learner corpora have made to the two disciplines of linguistics and second language acquisition, by referring to four electronic corpora which in some way are similar, but not identical to, the corpus described in this thesis. While Chapter 3 picks up references to the passive and places the construction more firmly in the forefront of the study, both theoretically and contextually, Chapter 4 will subsequently quantify existing learner corpora specifically in German, and discuss methods which are available for analysing them, in order to provide further evidence in exploring the development of the acquisition of a second language.

CHAPTER 3 - THE PASSIVE: RELATING GRAMMATICAL DESCRIPTION TO PEDAGOGICAL PRACTICE

3.1   Introduction

A brief introduction to the differences between the German passive and its English counterpart appeared in Chapter 1.3.   Having subsequently discussed theoretical descriptions of the passive with general application to a number of languages in Chapter 2, together with reviews of relevant SLA literature and learner corpora, Chapter 3 begins with descriptions of the passive which are used in the language teaching of L2 learners of German with L1 E. Grammatical definitions and descriptions of usage of the passive in both languages follow, together with comments on structures which exist only in German. It concludes with a discussion of a methodological tool used in the classroom teaching of the German passive.

In syntactic terms, the passive in all languages (insofar as it exists),[1] is indicated by using passive morphology of one type or another, whether it is the addition of a single morpheme 's' to the active form as in Swedish or an auxiliary verb combined with a past participle as is the case in English and in German. In semantic terms, the focus of meaning shifts in a passive clause from the subject of an active clause to the object. The syntax and semantics of the passive are introduced in the following sections.

3.2   The Passive in a Pedagogical Context

This section begins with definitions of the passive relating to English in reference works which indicate how the term is defined in grammar textbooks used by adult NS users. It is followed by a similar account of reference works which describe the German passive. This juxtaposition is chosen because any learners of an L2 have already acquired the grammatical structures of their L1 and bring that knowledge to their studies, whether or not they are aware of it. In particular, learners in an academic context may draw on their L1 when attempting to produce a grammatical structure in their L2, as has been pointed out in Chapter 2.5.3.9. We might expect that the L1 E learners of German L2 in this study would do this. These grammar textbooks would not only provide study

---

[1] As Keenan (1985: 248) points out, some languages do not have passives (e.g. the Sino-Tibetan Tamang and some languages in New Guinea).

material for a learner of German with L1 E seeking reliable examples of modern English and German, but would also serve for these learners as tools for the explicit learning of language which forms part of their learning experience, to accompany a particular learning method such as the one discussed in Section 3.6.   The implicit/explicit learning dichotomy has been explored in Chapter 2.

3.2.1   What Do Learners with L1 E Mean by 'the Passive'?

Although many adult users of English would not be able to define 'the passive' if they were asked, learners with an intellectual interest in language, and who are learning in the 'academic style' mentioned in Chapter 1, will be accustomed to referring to reference works if they were asked to define the construction. The definitions which follow come from Chalker & Weiner's dictionary (Chalker & Weiner 1998) and Greenbaum & Quirk's grammar (Greenbaum & Quirk 1995).

Chalker & Weiner define the adjective 'passive' as:

'Designating the VOICE[2] of the verb whereby the grammatical subject "suffers", "experiences" or "receives" the action of the verb'.

VOICE is then defined as:

A grammatical category which in English provides two different ways (ACTIVE and PASSIVE) of viewing the action of the verb. *Voice* is applicable to verbs, verb phrases, and entire clauses or sentences. The names *active* and *passive* are linked to meaning in that the subject of an active verb is often the actor or 'doer' of the verbal action (1994: 420).

The combination of these two definitions explains the importance of the verb in passive syntax, whether as a single word or as an item in a longer string of words building up to sentence level.

---

[2] SMALL CAPITALS used by the authors to indicate a cross-reference in the grammar dictionaries are retained.

Greenbaum & Quirk (1995) similarly draw attention to the differences between the passive and the active, but they include a reference to the semantic role changes in a clause as well as to the verb phrase. 'The difference between the active voice and the passive voice involves both the verb phrase and the clause as a whole' (Greenbaum & Quirk 1995: 44-45). In a passive clause in English,

a) [t]he active subject, if retained, becomes the passive AGENT[3]
b) [t]he active object becomes the passive subject
c) [t]he preposition *by* is inserted before the agent
(…)
(Greenbaum & Quirk 1995: 45).

As the extracts above demonstrate, Greenbaum & Quirk provide more linguistic information than Chalker & Weiner and since the formers' volume is specifically written for students it would typically be recommended as reference material to supplement knowledge of grammatical description in English, in addition to the two German grammar textbooks in English listed in Chapter 2.2.1. It not only supplies our definitions, it can reasonably be supposed that the definitions are known to the A cohort in this study. Chalker & Weiner's dictionary is aimed at a more general, non-specialist readership and is helpful both for its supplementary cross-references to the different passive types which will be mentioned in section 3.2.2 and the additional information from other theoretical perspectives which is useful but not imperative for this study e.g. indicating the meaning of 'verb phrase' in generative grammar which is not mentioned by Greenbaum & Quirk.

The 'action of the verb' (Chalker & Weiner 1998: 420), that is, the action or sentiment which the grammatical subject of the sentence performs or feels, is of particular significance in German. It is verb syntax and morphology in both German and English which directs the reader's or listener's interpretation of the meaning of the clause, but in German, case morphology and word order make it less simple than in English (where there is more or less no case morphology and the subject always precedes the verb) to identify the subject. Certain verbs must be followed by a specific case: *helfen* (help) is

---

[3] AGENT is defined as 'an animate which instigates or causes something. It is expressed by the preposition *by*, the complement of which frequently corresponds to the subject of a transitive verb' (Greenbaum & Quirk 1995: 200).

followed by dative case (DAT), whereas in English it would require a direct object, associated with accusative case (ACC). Word order in German clauses dictates that the verb appears in second sentential position (V2) but this changes according to clause function.[4]   Further, as in other languages, verb morphology helps to disambiguate constructions which might otherwise appear ambiguous. Compare, for example, (1) and (2) where (1) is a 3pers sing form and (2) a 3pers pl form. Both forms share a common pronoun *sie,* which can refer either to a single female or to many females or males (or a combination of the two).

(1)   *sie singt* (= she sings/is singing)
(2)   *sie singen* (= they sing/are singing)

The 'action of the verb' (Chalker & Weiner 1998: 420) thus affects not only phrase, clause and sentence structure but also meaning. Because both syntactic form and semantic meaning are implicated in the use of the German passive, the verb phrase has been chosen as the focus of this study. Chalker & Weiner's dictionary entry appears in the following terms: '(In Generative Grammar.)[5]   A sequence of words normally containing the verb and its complementation (i.e. virtually equivalent to the predicate), and forming, with the noun phrase which is its subject, a sentence' (Chalker & Weiner 1998: 417).[6]

In contrast to this generative description, for Greenbaum & Quirk, the term 'verb phrase' has two meanings: firstly it is applied to the verbal element in a clause as opposed to the subject or the object and secondly it is a 'member of a word class, like a noun and an adjective' (1994: 24). Hence a verb phrase may consist of one or more verbs (in the second sense), of which one may be an auxiliary verb (in the first sense). Greenbaum & Quirk's definition of the verb phrase is the one which is adopted in this study; the term as it is used here applies to the two verbs which form the German passive, an auxiliary verb (AUX) and a past participle (PP).[7]   This study focuses primarily on the auxiliary verb in German passive syntax.

---

[4] In generative grammar, German is considered a V-final language, with the basic position of the finite verb being reflected directly in the subordinate clause, not in the main clause structure. (Hawkins (1986: 135) gives a full description of German clause structure with reference to verbal word order).
[5] Capitals in the original.
[6] Chalker & Weiner's entry for *verb phrase* continues: 'See TREE DIAGRAM'. In Generative Grammar, this would show in diagrammatic form that a verb phrase may consist of more than just the syntactic verb.
[7] See also section 3.3.1.

(3)   The wasp stung the child.

(4)   The child was (AUX) stung (PP) by the wasp.

(5)   The wasp has stung the child.

(6)   The child has been (AUX) stung (PP) by the wasp.

In (4) and (6), which are the passive alternations of the active sentences (3) and (5), our attention is drawn to the child; in Chalker & Weiner's terms it 'suffers' being stung. The addition of a *by* phrase changes a simple passive to a complex one, in Keenan's terms (Keenan 1985).

3.2.2   The Passive in English Syntax

It has been explained in Chapter 2.2.2.2 that the English terms to be used to describe the two syntactic variants of the passive in German are *actional passives* and *statal passives*. According to Chalker & Weiner (1998: 371), statal passives 'express a state rather than a dynamic action' and are to be contrasted with actional passives, a term which is said to be 'the usual meaning of central passive verbs' (Chalker & Weiner 1998: 8). It is probably true that the majority of users of the English passive are not able to describe the semantic distinction between the two without some conscious reflection, since both types are based on *be*. L1 E learners in the A cohort, however, will have received some kind of explicit instruction and will know that the distinction exists in syntactic form in German. These learners will be able to draw consciously on their metalinguistic knowledge in order to produce correct grammatical utterances.

Chalker & Weiner (1998: 371) point out the ambiguity inherent in statal passives, when, for example, 'be' is used as a copula verb in combination with an adjectival phrase containing a PP to complete its passive sense (7).

(7)   The hall is decorated with flowers.

(7) has two interpretations, either that the hall is in the process of being decorated with flowers (an action is described) or it has already been decorated (a state). Greenbaum & Quirk also refer to this ambiguity and to the interpretation of a statal passive as an adjectival construction (Greenbaum & Quirk 1995: 45). If actional passives are the norm in English, it would not be surprising if these were to be acquired before statal

passives by learners with L1 E, but, as has been explained in Chapter 2.5.3.9, when learning German these learners have to disassociate actional passives from *be* and reassociate that verb in the other language with an alternative semantic form. Explaining that the *sein* passive could be viewed as an adjectival construction in German would, according to the description in the reference works, be a possible pedagogical solution in an explicit explanation.

3.2.2.1 Agent 'by' Phrases

While the passive voice changes the grammatical and semantic focus of a phrase, clause or sentence, it is also used when the intention is to deflect attention away from the grammatical subject of the active sentence. It may be that the agent is unknown or that the speaker or writer chooses not to mention it or indeed deflect attention away from it.

(8)   Peter Grubchick stole the money.

(9)   The money was stolen by Peter Grubchick.

(10)   The money was stolen.

In (4) and (6) the agent phrase 'by the wasp' may be excluded if, for example, it is not known whether a wasp or a bee stung the child, as Greenbaum & Quirk indicate. 'The prepositional phrase (AGENT *BY*-PHRASE) is an optional element and is commonly omitted' (Greenbaum & Quirk 1995: 45). In (10), no blame can be attributed to Peter Grubchick because the action of the verb is not attributed to him.

3.2.2.2 Verb Types

In English, only transitive verbs can be subject to a passive transformation e.g. (11) and (12). Greenbaum & Quirk state clearly that '[t]he distinction between active and passive applies only to sentences where the verb is transitive' (Greenbaum & Quirk 1990: 44). However, not all transitive verbs can be used with a passive sense. (9) shows the acceptable transformation of (8). (14), despite using a transitive verb 'lack' (13) would be disallowed as ungrammatical.

(11)   I deleted the file.

(12)   The file was deleted (by me).

(13)   My brother lacks courage.

(14)*   Courage is lacked by my brother.[8]

We shall see in section 3.3.1.1 that it is not only transitive verbs in German which can undergo a passive transformation.

3.2.2.3 'Get' Passives

The variety of means of expressing the passive in English has already been mentioned in Chapter 2. The numerous variations listed as Granger's seven-point typology are almost all rooted on the verb 'be', but there is an additional expression which does not require the same auxiliary. The 'get' passive (15) represents an alternative, more informal usage which would not commonly be found in academic language but which is semantically usually, but not always, identical to the 'be' passive (16).

(15)   Jessica Ennis got injured in training.

(16)   Jessica Ennis was injured in training.

Although it is not decisively clear in (15), we could deduce that the 'get' passive in English sometimes indicates a degree of responsibility of the subject with reference to the action of the verb, as indicated by Chalker & Weiner (1998: 172) and Greenbaum & Quirk (1995: 45). This particular sense of the 'get' passive may be seen in (17), where the suggestion is that Alfred was to a certain extent at fault in getting eaten. This is confirmed in the tale by the fact that he provoked the lion by poking it with his stick (Edgar 1932). (18) merely delivers a factual account of the outcome.

(17)   Alfred got eaten.

(18)   Alfred was eaten.

Use of the 'get' passive is often actively discouraged in academic discourse in English because of its informal connotations. We shall see in section 3.3.5 that the level of

---

[8] An asterisk preceding an example in this study denotes an ungrammatical expression.

informality attributed to the German equivalent of 'get', the verb *bekommen,* is the same.

In the next section we examine how an L1 E learner's pre-existing knowledge of the passive differs from the means of expressing it in German. The section begins with a description of passive typology in German, followed by a description of the formal structure of the syntax and associated variations of semantic function.

3.3   Defining the German Passive for Learners with L1 E

Duden (1998) provides the source of the categorisation and description of passive typology in German in this section. The Duden series of reference works has been the standard government-recommended description of the German language for the last 130 years, and is, in addition, used by students following DaF (*Deutsch als Fremdsprache =* German as a foreign language) courses, such as might be followed by learners with L1 E during a stay in a German-speaking country. The students in the A cohort may have used it during their time abroad and will be directed to it for reference purposes during their fourth year of study. Thus, both users of L1 G and learners of the language as L2 may use Duden as a point of reference if asked to define the passive and for that reason we use its descriptions of passive typology here. Although a comprehensive grammar of this type might be considered too complex for students at the level of those in the A cohort, its descriptions will reinforce the comparison between items which are typologically similar in both German and English, the languages which are L2 and L1 to the majority of the respondents in the A cohort.

At the university where this study was conducted, in addition to Duden, students who are less confident in the use of grammatical terminology are sometimes recommended to consult Zorach & Melin (1994) for the purpose of supplementing explicit grammatical knowledge because it is a grammar of both German and English.[9]   In addition to these two German grammar textbooks, the A cohort were also directed towards Durrell (1996) for additional clarification. The examples from these volumes which are given below would be familiar to the students in the A cohort.

---

[9] Note that the formation of the passive in this study guide is restricted to the use of *werden – sein* is not mentioned as a possible option for constructing the passive (Zorach & Melin 1994: 77). This point is discussed in Chapter 8.

In this discussion of the German passive, differences between the passive in German and English will be pointed out where these might indicate a problem for learners with L1 E. It is possible, for example, that these learners intuitively perceive the passive to be associated with 'be', a perception which would have to be amended with conscious reference to both the syntactic and semantic requirements of the passive in German, until the difference was fully acquired (if ever).

It was indicated in Chapter 2 that the perception of 'difficulty'[10] in acquiring the German passive may be linked to the low level of occurrence of the item in the teaching input (Hendrikson et al 2006). In the case of advanced level students, at least in their first year of study, explicit grammar instruction will provide additional focus on the construction and plentiful examples in context of both *werden* and *sein* passives would be included in order to promote acquisition. Despite this support, confusion may arise for learners with L1 E because of the variety of ways of expressing the passive in German, particularly where one of its types has no parallel in English.[11] Even for NSs of German, it would appear that the passive is not unproblematic. Eisenberg (2008) notes in his comments on Eroms' (2007) illustration of grammatical rules based on examples:

> Zum Standardpassiv stehen ihm sofort ein halbes Dutzend Alternativen zur Verfügung und er bemerkt, mit jeder von ihnen seien gleich auch einige grammatischen Klippen verbunden.
>
> (*He (Eroms) can choose from half a dozen alternative forms to illustrate the standard passive and he notes that there are grammatical pitfalls linked with every one of them.* My translation.)

3.3.1   The German Periphrastic Passive

The German periphrastic passive is formed by combining an auxiliary verb with passive morphology (in this case, PP), as described in Chapter 2.2.1. German verbs are classed as either regular (also known as weak) or irregular (also known as strong). The passive of most regular verbs is produced straightforwardly by combining a form of *werden* or

---

[10] That some learners perceive the passive, in particular the *sein* passive, to be 'difficult' is indicated by responses to a questionnaire which was completed by the contributors to the A corpus. The questionnaire appears as Appendix A. A full analysis is proposed for a future project.
[11] See Section 3.3.2.

sein with a PP which in turn is formed by replacing the final infinitival –*en* with –*t* and preposing the letters *ge*- (19).

(19) *decken* (= to cover) *wird/ist* *ge-deck-t*

Irregular verbs usually require the same prefix *ge*- together with a vowel change from the infinitival form (20).

(20) *stehlen* (= to steal) *wurde/war* *ge-stohlen*[12]

It was explained in Chapter 2.2.2.2 that the German terms to be used to describe the two variants of the passive in German are *werden* passives and *sein* passives. It will become evident that even with standard, basic passives, identified in Chapter 2.2.1, learners with L1 E have a choice to make when producing the construction. That is not to say that learners make a conscious choice. Their prior exposure to passive forms, combined with their pre-existing L1 knowledge, will presumably enable them to comprehend the passive before they are able to produce it. Nevertheless, the syntactic differential does not exist in English and therefore presents the L1 E learner of German with an additional item to process. The decision-making process becomes even more complex if 'get' passives, outlined in Section 3.2.2.3 as sometimes, but not always, similar in meaning to 'be' passives depending on the degree of responsibility of the grammatical subject, enter the equation.

However, although the meaning of the *werden* passive can often be interpreted as the informal 'get' passive, the focus here is not on the variation in the level of formality in English but on the difference in usage between the *werden* and the *sein* passive in German and the degree to which learners with L1 E produce it in a NS-like way.

Assuming that the learners have been exposed to a certain amount of explicit instruction on how the passive is formed, in order to produce a semantically accurate passive they must either be capable of accessing that explicit knowledge and have time to do so, or

---

[12] Variations on the construction of PP include verbs with separable prefixes e.g. *zusammenfassen* (regular verb = to summarise; PP *zusammengefasst*): from *zusammen* (together) and *fassen* (hold or bring) and those with prefix *ver* - e.g. *verlieren* (irregular verb = to lose; PP *verloren*). These examples indicate the number of variables involved in the construction of the German passive.

they will have acquired the difference between the two types and will know intuitively, without conscious reflection, whether to use *werden* or *sein*. The morpho-syntactic difference between *werden* passives and *sein* passives in German is self-evident in (21a) and (22a). The difference between the two in English, however, (21c and 22c), is less obvious since the verb form *is* suffices for the interpretation of both types of German passive.

(21) a) Das Pferd wird an den Baum gebunden.

  b) The horse is being/becoming to the tree tied.

  c) The horse is (in the process of being) tied to the tree.

(22) a) Das Pferd ist an den Baum gebunden.

  b) The horse is to the tree tied.

  c) The horse is tied to the tree. (The action is completed)

  Durrell (1996: 299)

Even when the choice of auxiliary has been established, the requirements of the German verb in terms of respecting the case system must be considered. The focus of this study, as stated earlier in this chapter, is the verb phrase used to express the German passive, because NS-like representation of the construction depends on both syntactic and semantic knowledge which differs from English and is therefore of particular interest in a pedagogical context where learners aim for accuracy in order to achieve study goals, as indicated in Chapter 1.4. The next section looks at verb types in connection with the construction of the passive, since verb typology influences case morphology and increases the complexity of the process of producing a passive expression. Learners must simultaneously consider not only verb morphology but also associated changes to case endings on adjectives and nouns.[13]

3.3.1.1 Verb Types

Duden (1998: 172) categorises three types of *werden* passive: A (a transitive verb requiring a direct object in the Accusative case e.g. *verkaufen* = to sell), B (an

---

[13] The study does not consider the acquisition of case morphology of associated subject/object noun phrases.

intransitive verb followed by Genitive case[14] (e.g. *gedenken* = to commemorate (elevated language), Dative case (e.g. *helfen* = to help) or a prepositional object e.g. *denken an etwas* = to think of something) and C (intransitive verbs with no object e.g. *tanzen* = to dance). Their rates of occurrence are 97%, 2% and 1% of finite verbs respectively, indicating that an actional passive with a direct object (Duden type A) is by far the most common passive found in Duden's selection of contemporary German texts, whereas the occurrence of Type C is minimal, both by comparison with the numbers of other passive types and with passive occurrence overall.[15]

(23) Type A

Das Buch   wird   verkauft.

The book   becomes   sold.

The book is/is being sold.

(24) Type B

Ihm   wird   geholfen.

Him (DAT)   becomes   helped.

He is being helped.

(25) Type C

Es   wird   getanzt.

It   becomes   danced.

People are dancing.

As was noted in section 3.2.2.2, only transitive verbs can be used in a passive form in English, so learners with L1 E may find types B and C more difficult to interpret or to produce.[16]

A further point in the consideration of verb types is that certain Type A transitive verbs in German cannot be used to form a *werden* passive, for example those which indicate

---

[14] Verbs followed by the Genitive case are rare. The inclusion of them by Duden indicates the breadth and depth of grammatical description in this volume.
[15] The Duden corpus is regularly updated through online searches of a number of sources, for example newspapers, magazines, novels and repair manuals. It currently numbers more than two billion word forms (Duden online 2014).
[16] See (14) for an example of a transitive verb which cannot appear in the passive in English.

possession, but the disallowed ungrammaticality also exists in English (26). It is often the case that an active form is the acceptable equivalent (27). Thus there are cases where passives in the two languages do not show variation but these are few in number and constitute a further consideration for an L1 E learner of German who intends to produce a passive expression.

(26)   *Vier Computer   werden   von meiner Kusine   gehabt.
        Four computers   become   of my female cousin   had.
        *Four computers are had by my cousin.

(27)   Meine Kusine   hat   vier Computer.
        My (female) cousin   has   four computers.

3.3.2   Subjectless Passives

In contrast to the similarity discussed in the previous paragraph, a Duden Type C verb, which in English is termed a 'subjectless passive' by Durrell (1996: 296) and 'impersonal' by Zorach & Melin (1994: 210), can be used when there is no known agent mentioned. According to Durrell, 'no comparable construction exists in English' (Durrell 1998: 297). In a subjectless passive, either the pronoun *es* (= it) is inserted in sentence initial position if there is no other element pre-posing the verb (25) or it is omitted all together (28).

(28)   Dann   wurde   auf den Straßen   getanzt.
        Then   was (no subject)   in the streets   danced.
        Then there was dancing in the streets.
        (Durrell 1996: 297)

3.3.3   Agent 'by' Phrases

Agent 'by' phrases in English were described in Section 3.2.2.1. In German, the agent in a passive clause is expressed by the addition of the prepositions *durch* (= through/by) or *von* (= from/by) to indicate the actor of the verb, demonstrated in (30), which is the passive variant of the active (29). The agent phrase may be omitted in German as well as in English (31). The structure preposition + noun is similar in both English and

German, but, as indicated in section 3.2.1 the focus in this study is on the passive verb. Agent phrases are included in the description of passive formation and usage in order to give a comprehensive picture of the complex procedures a learner may have to negotiate in producing the passive in German.

(29)   Peter (NOM)   schreibt   den Brief (ACC).
       Peter   writes/is writing   the letter.

(30)   Der Brief (NOM)   wird   von Peter   geschrieben.
       The letter   becomes/is becoming   by Peter   written.
       The letter is being written by Peter.

(31)   Der Brief (NOM)   wird   geschrieben.
       The letter   becomes/is becoming   written.
       The letter is being written.

3.3.4   Word Order and Sentence Structure

Not only is the verb typology complex in German, the associated sentence structure and case system combined represent a large number of variables which the learner must manipulate simultaneously before producing an utterance, as suggested in Section 3.3.1.[17] When the passive in English, e.g. (4) and (6), requires the object noun phrase in the equivalent active clause to be in initial position in front of the verb, e.g. (3) and (5), then it becomes the subject of the clause. In semantic terms, in a passive sentence in English, the focus falls on the new information which appears after the verb e.g. (4) and (6). This is not necessarily the case in German, where semantic roles can be expressed syntactically by case marking and not just by sentence position. As long as the verb is placed second (final in a subordinate clause), the grammatical subject can occupy post-verbal position even in an active sentence if the initial position is occupied by an adverbial, for example (32).

(32)   Gestern   schrieb   Peter   den Brief (ACC).
       Yesterday   wrote   Peter   the letter.
       Peter wrote the letter yesterday.

---

[17] See also Hawkins (1984) and Duden (1998).

(33) Gestern   wurde   der Brief (NOM)   von   Peter   geschrieben.

Yesterday   became/was becoming   the letter   by   Peter   written.

The letter was written by Peter yesterday.

(34) Gestern   wurde   der Brief (NOM)   geschrieben.

Yesterday   became/was becoming   the letter   written.

The letter was written yesterday.

In the passive sentences (33) and (34), it is case marking, not position, which determines meaning.

3.3.5   The *bekommen/kriegen* Passive

In Section 3.2.2.3, the use of the English auxiliary 'get' was described as a largely informal alternative to the verb 'be'. A similar construction is possible in German, although without the element of responsibility of the agent attached to it. Students writing academic German would be advised to respect a formal register and to avoid expressions which are suitable only in informal settings. To exemplify this point, we refer to Lühr (2010: 172), who cites Vogel's study of German chat-room language (Vogel 2003) in which the following examples of both the *bekommen-Passiv* (35) (*bekommen* = to receive, get) and the *kriegen-Passiv* (36) (*kriegen* = to get (even more informal than *bekommen* (Duden 1998: 181)) occur:

(35) Hab   neulich   ein Buch   geschenkt   bekommen.

(I) have   recently   a book   presented   received.

Gotta book as a present recently.

(36) Ich dachte,   das   kriegt   man   mit windows   geliefert.

I thought   that   gets   one   with windows delivered.

I thought that comes with Windows/ I thought you got that with Windows.

Because it is considered an avoidable means of expression, and is actively discouraged in formal German as well as in English, the *bekommen/kriegen* passive is not expected

to occur in the texts in either the A or the B corpus. Even if L1 E users of the German passive are not able to produce the full range of options in expressing the passive, it can be assumed that they will be aware that low-register expressions such as *bekommen/kriegen* should not be used in academic discourse. This construction has therefore not been included in the concordance search because the study is primarily focused on the learners' selection of auxiliary verbs in the German passive rather than pragmatic use of the passive in a variety of contexts.

The next section contains a description of alternative ways of expressing the passive which are available to learners of German but which may not have straightforward equivalents in English.

3.4   Alternatives to the Passive in German

Alternatives to the two 'standard' periphrastic passives (AUX + PP), a reflexive (*sich lassen* = to allow something to be done), (37) and use of the indefinite pronoun *man* (= one; someone; we) with an active verb (38) are explained in Duden (1998: 173-182).

(37)   (a) Das Fenster   lässt   sich   leicht öffnen.
         (b) The window   allows itself   easily to open.
         (c) The window can easily be opened.

(38)   Man   kann   das Fenster   leicht öffnen.
         (Someone)   can   the window   easily   to open.
         The window can easily be opened.

The type of reflexive in (37b) would not be possible in English because a window cannot open itself.[18]

The indefinite pronoun *man* (= someone, one, we) can sometimes be used as a substitute in German for a passive intention. However, since it replaces a noun + determiner (DET) (in generative terms, a noun phrase) an agent, if expressed in a passive sentence,

---

[18] The alternative form 'The window opens easily' does exist, however. In this case, passive meaning is suggested but it is expressed in an active form.

would be omitted if an active version was attempted. Hence the sense of (25) is similar to (39) but (40) loses the reference to Peter which is included in (29).

(39) Man tanzt.

People dance/are dancing.

(40) Man schreibt den Brief.

Someone writes/is writing the letter.

Or

The letter is written.

A third alternative to the periphrastic passive in German is described in Duden (1998: 181). In the section on alternative passive typology, Duden includes the verb *sein* (be)+ *zu* (to) + infinitive (INFIN) but yet another layer of potential confusion for any learner is couched in the following caveat:

Das Zustandspassiv kann leicht mit anderen Konstruktionen verwechselt werden, deren Prädikatsverband - der äußeren Form nach - ebenfalls aus sein + 2 Partizip besteht.

The statal passive can easily be confused with other, similar-looking constructions whose predicate is formed with sein and PP.

(Duden 1998: 181).

This alternative to the *werden* passive is mentioned in the qualitative discussion because, as has been said, the English passive is based on 'be' and any construction including this verb might present fewer problems in production (41) for L1 E learners. This alternative equates to another possible passive variant: a modal verb requiring INFIN to complete its sense, but in this case the auxiliary *werden* is required (42).

(41) Das Fenster ist leicht zu öffnen.

The window is easily to open.

The window can easily be opened.

(42) Das Fenster kann leicht geöffnet werden.

The window can easily opened to become

The window can easily be opened.

Other alternatives to the actional or statal passives in Duden's typology are not included in this comparison between German and English as they are unlikely to occur with any regularity in the instructional input to the L1 E learners in the A cohort, e.g. *gehören* (= belong) + PP, which Duden considers colloquial and confined to southern German-speaking areas (Duden 1998: 181).

Although the stated focus in this study is the verb phrase as defined by Greenbaum & Quirk (1994: 24), the value of these explanations of changes in sentence structure and syntax is to support the claim that the complexity of the German passive poses additional problems for learners, particularly those whose mother tongue equates the passive with the verb 'be'.

3.5   Frequency of the German Passive Construction

According to Duden (1998: 172-173), only 7% of finite verb forms (*in Texten der deutschen Gegenwartssprache* = in contemporary German texts) are passives, of which 5% are actional passives and 2% statal passives. In comparison, instead of measuring passive tokens (i.e. examples of the passive expressed by AUX + PP) against the number of finite verb forms, passive incidence in the British National Corpus of Written English is measured against the number of words in the text in question at approximately 8 per 1000 (Biber 1988).[19] On a percentage basis, this results in a token count of less than 1%. Based on these figures from German and from English, low frequency totals are to be expected in the corpus collected here, which has been produced by learners with an existing knowledge of descriptions of the passive in English who encounter a variety of constructions which perform the same function in German. In this study we represent the frequency of the passive as the number of passive tokens per-500-words. The term *tokens* (individual items) encompasses the different aspectual and morphological *types* of passive forms, the significance of which is described in Chapter 5.

---

[19] 'be' and 'get' passives.

3.6 Teaching the German Passive: an Instructional Example

The DaF course '*Begegnungen*' (Buscha & Szita 2007) is chosen as an example of pedagogical materials which are used in the instruction of L1 E learners of German at university level. While the L1 E learners in this study will have received tuition based on an English teaching method at secondary school, in the university where this study took place, FL textbooks using the target language throughout are favoured for the teaching of beginner learners in order to support as rapid acquisition as possible. This strategy accords with the communicative and usage-based language teaching approaches in which the effects of the learning activities are maximised through 'the (exclusive) use of the target language by the teacher' (Whong 2011: 133). The 3-volume *Begegnungen* series is comparable to an English-based beginners' textbook. *Begegnungen A2* is the second in the series leading to the next (B) level of the Common European Framework of Reference for Languages (CEFR). The passive appears for the first time in Chapter 8 of 8 in this volume, making it the final grammatical element to be introduced in this section of level A.[20] Notably, it is the *werden* passive which is introduced in the 3pers; the *sein* passive is not mentioned, suggesting that the focus of attention is the action of the passive rather than the state which results from it. Further evidence of the focus at this early stage of instruction may be identified by the fact that exercises are accompanied by the statement '*Bei einem Passivsatz steht die Handlung im Vordergrund, nicht die Person*' (= In a passive sentence it is the action which is foregrounded, not the person) (Buscha & Szita 2007: 234). In its earliest iteration there is no mention in this description of the passive of an agent 'by' phrase. Such a description supports our choice of Chalker & Weiner's definition of the passive as offering a focus on the verb for the initial stages of learning the German passive for learners with L1 E.

3.7 Review of Chapter 3

In strictly formal terms, it can be seen that the German passive presents learners with an enormously complex and varied set of options. Because the construction appears with such low frequency, as has been shown in Section 3.5, there will be sparse evidence of

---

[20] While the CEFR level A1 expects a learner to be able to connect simple phrases and sentences, level B1 learners aim for more detailed texts 'on a variety of subjects related to his/her field of interest, synthesizing and evaluating information and arguments from a number of sources' (European Council 2014).

this complexity and variety in the linguistic input, although it will be more frequent in situations where input is enhanced by explicit instruction of some kind, such as PI as described in Chapter 2.5.3.1. For learners with L1 E, then, for whom the passive is formed with the verb 'be' + PP and is limited to transitive verbs, and who, according to Duden's calculation (Duden 1998: 172) will be exposed to only a limited number of occurrences of the passive in the input, it is perhaps not surprising that the complexities of expressing the passive in German elude all but those learners whose mastery and control of the language is at a very high level. In addition, learners may not gain all that is intended from the use of grammar textbooks, since the language they portray may not be an exact replication of everyday usage but may only display all the possible variations of a particular construction. While Durrell does say that the passive is used less frequently in German than in English (Durrell 1996: 294), nothing more specific is identified, whereas Duden delineates the percentage difference of occurrence of *werden* and *sein* passives in German (Duden 1998: 172), and Duden online (2014) alludes to the types of contemporary German texts on which the volume has based its calculation (see footnote in Section 3.3.1.1).

The intention in Chapter 3 was to describe the differences in the ways of expressing the passive in English and in German and to indicate how these differences would be conveyed to adult learners of German with L1 English. The three types of German *werden* passive, plus the semantically different *sein* passive and the further addition of three alternatives to the passive present a learner of German with a number of options which do not exist in English. It is probably not the case that a learner consciously debates whether one or the other form of the passive will be used or whether it can be avoided each time that he or she decides a passive form is required. What is more likely is that the most easily accessible form in the learner's grammatical system will be used, and this form may be the one which occurs most often in teaching input or in naturalistic input outside the classroom. Nevertheless, the number of alternatives would offer one explanation for low frequency levels of the periphrastic passive in the interlanguage of advanced learners, which were mentioned in Chapter 2 in connection with input in a formal learning environment and will be considered again in Chapter 6. Despite the low frequency counts in contemporary language (Duden 1998: 172-173), however, some evidence of other types will be expected in the study of advanced learner interlanguage because of the expectation that learners at this level will show some elements of more mature, if not necessarily NS-like language.

The hierarchy of passive forms that will be proposed in this study and which might therefore be expected to appear in the interlanguage of learners of L2 German at this level is based on the examples in their written production of simple *werden* and *sein* passives as described in the sections above. If, as research has indicated (e.g. Abbot-Smith 2003), L1 learners of German acquire the *sein* passive before the *werden* passive, and developmental sequences of other linguistic structures have been found to be similar in L1 and L2, but the interlanguage of the L2 G learners demonstrates the opposite, might it be the case that instruction in the case of the passive in this particular context does in fact affect the order of acquisition? Is this evidence of the effect of pedagogical input?

Answers to these questions may emerge from an examination of the data which have been collected. In Chapter 4, returning to the research questions posed in Chapter 1, a suitable method of investigation is identified. A number of electronic tools used to analyse learner language are reviewed and the subject of annotation is discussed. The advantages of a mixed-methods analysis will also be presented in Chapter 4, and having determined the most appropriate way of eliciting second-stage data from the data collection, the application of the selected methodology to the LINCS corpus will be described. This will lead in Chapters 5, 6 and 7 to a detailed analysis of the data on the passive which are extracted from the corpus and thence to reflections on the acquisition and use of the syntactic forms of the German passive by learners with L1 E.

CHAPTER 4 – IDENTIFYING AND IMPLEMENTING A METHODOLOGY

4.1   Introduction

In consideration of research aim 2, to offer a new corpus to a wider research public, it is useful to reflect on the principles underlying the construction of a learner corpus, in order to justify the choice of this method of data collection and analysis in the study reported here. In order to do this, and to outline the expected contribution of the present study to the fields of linguistics and SLA with examples from a learner corpus, the intention in the first section of Chapter 4 is to survey a number of corpora of learner German. This will refine the focus of this study from the more general review of the limited number of studies of the German passive which was carried out in Chapter 2. The description in that same chapter of three corpora containing languages other than German and one which includes a limited amount of German data served as an introductory overview of the development of the genre.

Initially, the current spectrum covered by shared databases will be pointed out. Secondly, a comparison will be made between existing corpora and the LINCS corpus. In the next section of this chapter, a number of tools which are available for analysing corpora are described and considered in terms of their usefulness in relation to this study. The research aims are restated for clarity and the term 'passive intent' is introduced in order to support the choice of a concordancing program as the most useful tool in this context, and the benefits of implementing electronic means of analysis are restated. Some limitations inherent in the application of a purely technology-based investigation are also considered, together with the proposal that the effect of any limiting factors may be reduced by the implementation of a mixed quantitative/qualitative, cross-sectional/longitudinal methodology.

Having defined the preferred methodological approach to the study, the composition of the LINCS corpus is described, followed by the precise methodological procedures which are implemented here, beginning with the preparation for analysis, and subsequently elaborating the procedures for the preliminary study, the quantitative and finally the qualitative investigation. Mention is made of further data which are available

for future processing in order to contribute an additional dimension to the results of this study.

4.2   Data Sharing in Learner Corpora

If the possibility of storing large collections of data electronically means that retrieval becomes simpler, a further advantage of this type of collection lies in the facility of data sharing. As Adolphs says, 'Electronic text analysis lends itself to collaborative research, as most texts can now be exchanged relatively easily between users via file transfer protocols' (Adolphs 2006: 28). Electronic collections may consist of either oral or written data, depending on the research aim. As indicated in Chapter 2.5.3.4, written data has its place in studies with a particular focus on language pedagogy, such as the one described here, where it is possible to compare learners' production during timed and non-timed exercises. The methodology adopted in this case, a version of CIA as described in Chapter 2.6.2, compares the IL of learners firstly with the same L1 on the same task and secondly with NS language.[1] It is intended that the LINCS corpus will, in time, be added to an existing corpus in order to increase the amount of data available for future SLA research. Because the ICLE (introduced in Chapter 2.6.3.2) was one of the first learner corpora to be developed and is maintaining its reputation through a continuous process of development, it is used as a benchmark here in the comparative discussion which follows.

The ICLE website describes not only its own corpus of learner English, covering 16 mother tongue backgrounds in 6,085 essays and numbering 3.7 million words, available to purchase on CD, its subsection of learner corpora around the world lists a further 124 corpora, the majority of which are freely available or are likely to be at a later date, following full development (Université Catholique de Louvain 2013). Several of these 124 number more than one million words and most of those collections are corpora of L2 English. It is easy to see where the balance of current collections lies in terms of target language (TL).

---

[1] CIA may also contrast the FL produced by learners with different L1s, as indicated in Chapter 2.6.2. This would be possible if the LINCS corpus were expanded to include the German production of learners with, for example, L1 French or L1 Spanish.

In the UCL list, it is noteworthy that only nine corpora have German as the TL, including the LINCS corpus. One of the remaining eight (excluding the LINCS corpus) is not publicly available, one focuses on spoken data rather than written and a third is also excluded from the comparison which follows because the L1 of the learners is Italian. Five corpora can thus be considered in more or less comparative terms with the corpus which has been collected for this study, that is, they all contain written essays in German. These five corpora are schematised in Table 4.1. A blank segment indicates that the information is not available for this item.[2] For comparative purposes, outline details of the current constituents of the corpus described here (the LINCS corpus) appear in Table 4.2.[3] The A corpus contains the German language of the A cohort in the form of essays of varying lengths; the second parallel subsection of the corpus, the B corpus, contains a number of FL texts written by students with German L1 and the C corpus contains text input used in advanced German language classes.

4.3   A Comparison of Existing Learner Corpora with the LINCS Corpus

For the purpose of making an exact comparison with the LINCS corpus and any of the five in Table 4.1, identical statements of L1 and L2, content, purpose and level are required. The availability of the corpora, or the means of access to them, will also be considered.

Of the five corpora listed in Table 4.1, the AleSKo corpus would seem to be similar to the LINCS corpus by virtue of the fact that the TL is German and the task type is similar at least to the Year 4 exercise reproduced in the LINCS A corpus, full details of which appear later in this chapter.[4] However, it is not used as a comparison here because the learners' L1 is Chinese as opposed to English and the German content of the corpus comprises L1 language rather than L2. In addition, no information is given concerning the Proficiency Level and Size in words of the AleSKo corpus.

---

[2] Note that the UCL list pertains only to learner corpora. Among other corpus collections of a different type is the publicly available collection of German newspaper language which has been assembled in Tübingen.
[3] A full description of the LINCS corpus follows in Section 4.6.
[4] See Section 4.6.1.1.

Corpus	Target	First language	Medium	Text type/task type	Proficiency level	Size in words	Project director	Availability
The AleSKO corpus	German	Chinese (but also German L1 data from the FALKO corpus)	written	Argumentative essays				
Analyzing Discourse Strategies: A Computer Learner Corpus	German	English (mainly American English)	written	Threaded Discussion Chat Essays Longitudinal	From beginner to intermediate-mid	under development	Christina Frei Edward Nixon University of Pennsylvania	
The Corpus of Learner German (CLEG13)	German	English	written	Argumentative, free compositions Longitudinal over 4 years, undergraduate students	Intermediate to advanced	c.320,000	Ursula Maden-Weinberger	Online access through the FALKO platform. Please contact Ursula Maden-Weinberger
The FALKO corpus (Fehlerannotiertes Lernerkorpus 'error annotated learner corpus')	German	various	written	1. Falko summaries 2. Falko essays 3. Falko Georgetown: letters, fiction writing, journal articles, book reviews (= longitudinal data from American learners)	Mainly advanced (Falko Georgetown: beginners – advanced)	Summaries: 41,072 Essays: 23,579 Georgetown: 126,105	Anke Lüdeling Maik Walter Humboldt-Universität zu Berlin Institut für deutsche Sprache und Linguistik, Germany falko-korpus@hu-berlin.de	Online access
The LeKo (Lernerkorpus) corpus	German						Anke Lüdeling, Humboldt-Universität Berlin, Germany	Online access (password protected) Register here

Table 4.1: Extract from the UCL list of learner corpora around the world
Reproduced with permission from http://www.uclouvain.be/en-cecl-lcWorld.html (Last accessed 15-12-13)

Corpus	Target language	First language	Medium	Text type/task type	Proficiency level	Size in words	Project director	Availability
The LINCS corpus	A. German	A. English	A. Written	A. Essays; examination answers. Longitudinal and cross-sectional data	A. Intermediate to advanced	c84,180	Elizabeth Thoday	Not currently publicly available
	B. German	B. German	B. Written	B. Essays	B. Advanced			
	C. German		C. Written	C. Teaching input				

Table 4.2: The LINCS corpus (table based on UCL at http://www.uclouvain.be/en-cecl-lcWorld)

The Analyzing Discourse Strategies learner corpus is similarly disregarded, primarily because the proficiency level of writer competence is lower than that in the LINCS corpus, with the longitudinal range starting at Beginners' level and progressing to Intermediate. Nevertheless, it is one of few longitudinal studies of learner language,[5] and might be worthy of further investigation if the LINCS corpus were to be used in the future to analyse discourse from a functional perspective, for example.

The 320,000 words of the Corpus of Learner German (CLEG13), on the other hand, might serve as a useful comparison with the LINCS corpus which is still in an embryonic form. However, it has not been considered in this study because it has only recently been made publicly available. At the time that the study described here commenced, Ursula Weinberger's thesis was only available on personal request and the tagging system remained under her personal jurisdiction. At the time of consulting that thesis, error tagging was still under consideration as a possible treatment for the LINCS A and B[6] corpora but it was subsequently disregarded for the reasons which are elaborated in Section 4.4.1. As can be seen in Table 4.1, the CLEG13 corpus is now available through the **F**ehler**a**nnotiertes **L**erner**ko**rpus (FALKO) platform.

Both FALKO and LeKo (**Le**rner**ko**rpus = *learner corpus*), are under the direction of Anke Lüdeling (Lüdeling 2006) who indicated that LeKo 'is not maintained and documented - it cannot be used for any kind of serious research' (Lüdeling 2010). LeKo is therefore disregarded in this comparison and it is only the FALKO corpus with which we draw a comparison, based on Table 4.1. An explanation of the purpose of FALKO will serve to underline the contribution which the present corpus, the LINCS corpus, is expected to make to the field of learner corpora. It will be considered with respect to the error-tagging system under which the data are stored, in connection with the alternatives available to users wishing to carry out parameter-specific examinations of language.

---

[5] A call for an increase in the number of longitudinal studies has been made throughout the history of SLA by, among others, Corder (1981) and Myles (2005).
[6] See Section 4.6.1.2.

4.4   Analytical Tools

Granger (1994) ascribed the upturn in the use of corpora to the availability not only of computerised collections of data but also of the tools with which to analyse them. Smith, Hoffman & Rayson (2008) indicated the extent of advances in corpus tools, while also specifying the need for collaboration between linguists and software developers to ensure that the needs of the former are met by the availability of suitable technology. Scott (2010) describes the rapid development not only of his own analytical toolkit but also of computer capabilities in general. He provides an overview of the possibilities afforded by the advances in technology specifically in producing frequency counts, which is of particular relevance to this study. Error taggers, part-of-speech (PoS) taggers and concordancers are described here, together with an assessment of their usefulness in this particular investigation. Support for the approach which is taken in this study and explained in this section, i.e. the analysis of a non-annotated corpus, is to be found in Evison's (2010) exploration of basic corpus analysis techniques.

4.4.1   Error Taggers

FALKO has been developed considerably since it first became publicly available in 2005 (Lüdeling et al 2005) and two further ontological systems were subsequently added to it (Chiarcos et al 2008) enabling multi-level comparison of data from different, pre-tagged systems to be compared electronically. However, as Rastelli (2009) points out, error tagging is just that – all you discover are the errors, without any further reference to the way in which these errors were produced. Since the identification of an error is likely to be subjective, in some way tied to what is deemed by the teacher or the researcher as correct, and usually bound to some norm of nativeness, it would seem that error tagging alone will not offer the evidence for the development of a linguistic structure that is being sought in this project. An error tag might reveal morphological discrepancy, for example, but this type of information was not required for the study described here. An additional, manual examination of the data would be necessary in any case in order to identify passives and comment on their usage.

Limiting the assessment of an utterance to only two possible outcomes, correct or incorrect allows only a broadly simple result, what Chiarcos et al (2008) term 1: n-

mapping or 1:1 mapping. Their article describes the ANNotation of Information Structure (ANNIS) tool, which is designed to address this issue and offer a multi-level system of analysis. Zeldes et al (2009) further elaborate the ANNIS system, suggesting that it enables multi-layer corpora to be created by more than one annotater for the same section of a corpus. The tool also allows annotated texts to be searched and the results to be displayed in a number of ways, depending on the research query: in a grid view, examining non-hierarchical data (such as tokens in learner language), in a tree view (for syntax) and a discourse structure view (for examining contextual information). Zeldes et al (2009) indicate how the FALKO corpus may be queried in CIA using the ANNIS tool. A similar multi-level analysis can also be achieved manually, by individuals, but this is enormously time-consuming without guaranteeing perfect results. While these methods undoubtedly address some of the possible discrepancies inherent in the subjective tagging by individual interraters, even such advances do not provide the finer type of qualitative assessment which is required when studying a learner's interlanguage. This was an additional reason why, in the course of evaluating analytical systems, annotation seemed unnecessary for the study described here.

4.4.2   Part-of-Speech Taggers

Part-of-Speech (PoS) taggers, as the name suggests, 'tag' or 'mark' words within a text according to a pre-arranged set of types or, in grammatical terms, parts of speech. In this way, a tagger identifies specific grammatical elements at a number of different levels, whether syntactic, morphological or lexical. A second stage in the annotation process conducted by a PoS tagger produces a skeletal analysis by dividing the text into bracketed subsections according to syntactic categories such as adverb, verb or noun phrases e.g. the Constituent Likelihood Automatic Word-tagging System (CLAWS) program at Lancaster University (Garside 1987) and the Penn Treebank (Marcus, Santorini & Marcinkiewicz 1993). Systems such as these were initially 'trained on native speaker English' (Pravec 2002: 99) and were therefore unsuitable for use on corpora of other languages because only the English lexicon is recognised. A small number of PoS annotation systems exist for German, for example the tagger developed by the Sketch Engine design team, taking its data from the web, which is aligned to the design of tools for Machine Translation (Schmid & Laws 2008), the CHILDES CHAT and CLAN tools (see Chapter 2.6.3.2) and the **M**orphologisches **Pro**gramm (MPRO)

(Maas 1998) the last of which forms the basis for ongoing research by Paul Schmidt in Saarbrücken.[7]

4.4.3   Concordancers

A concordancer sorts data according to parameters specified by the user. Its principal function is to search a body of text and identify each example of a particular token (i.e. item) or tokens, listing them enclosed within a pre-selected number of words as illustrated in Section 4.8.2.1. The total number of tokens retrieved by the program can then be set against the pre-selected word to arrive at a token/word ratio. A concordancer cannot retrieve specific parts of speech; a PoS tagger is required for this latter procedure. Based on an appraisal of the three types of electronic tool mentioned here, it seemed that none of them individually would provide a complete response to the research questions posed in this thesis because a qualitative comparative assessment of language is required in addition to a straightforward count and this type of qualitative comparison cannot be carried out by an automated instrument.[8] A method of identifying examples of the two constituent parts of the German syntactic passive was required together with a measure of the users' decisions as to the context in which they produced *werden* and *sein* passives.

4.4.4   To Annotate or Not To Annotate?

In the time which has elapsed even since the commencement of this project, the power and flexibility of the available tagging systems have increased. In accordance with research goal 2, the LINCS corpus in its original format will in due course be offered for tagging by one of these automated systems, in order for the data to be used in future research studies, but at the pre-analytical stage use of a tagging system did not appear to offer a suitable means of sorting the data. At the later, full analysis stage, the decision was taken not to annotate the corpus as it currently exists, since it is relatively simple to manipulate manually a limited number of words without annotation and multi-level tags are unnecessary when the focus is on only one structure.

---

[7] A fuller discussion of MPRO appears following Figure 4.1.
[8] The way in which the term 'comparative' is used in this context is specifically related to Granger's (2004: 127-128) description of the four main functions of learner corpus analysis, which are explained in Section 4.4.4, and to the process of CIA, elaborated in Chapter 2.6.2.

Although many collections of learner language now exist, as explained in Section 4.3, and annotation systems have become increasingly sophisticated and are capable of carrying out rapid tagging procedures, an untagged corpus remains the evidence base for this project which also allows ease of access for anyone wishing to replicate the study or use the data for subsequent research. It should be noted that the latest version of the ICLE not only includes a built-in concordancer, the original texts are also retained so that individual researchers may choose an analytical tool to suit their own intent and purpose.

In general terms, supporters of annotation regard it, according to Sinclair, as 'indispensable'; those researchers who base their findings on intuitive (i.e. corpus) data see annotation as 'obfuscating' (Sinclair 2004: 56). Putting the case against annotation, Sinclair had remarked that 'a particular view of language is imposed on the corpus, [by annotation; my clarification] down to the finest detail, and it is non-negotiable' (Sinclair 2004: 54). He advances the argument that the majority of researchers who use ready-annotated texts pay scant attention to the methods that have been used to prepare them, and that many researchers require only a fraction of the data which is recordable by annotation. Figure 4.1 indicates the many levels at which a single item can be categorised by MPRO, which assigns many levels of linguistic information to every word produced in an utterance.[9]

{ori=wird,lu=werden,mlu=werden,snr=163,wnra=3302,wnrr=20,offset=19712,ew= 0,lw=yes,last=no,pctr=yes,pctl=no,c=verb,vtyp=fiv,hsns=ns;hs,nb=sg,per=3,tns=pres,gra=small,w=1,cs=v,ds=werden,ts=wird,t=werden,ls=werden,ss=c0c4;c0c8,endung=0,s=c0c4;c0c8,lng=germ,mwa=nil,rco=0,lco=0,zf=werden,gram~=4121de;4122de,mark=vg_fiv,markcl=ns;vl;fiv;nf,chunk=12;13;14;15;16;17;18;19;20,head=20}

Figure 4.1: An automated tag of *wird* using MPRO

Such complex results of an electronic analysis provide much more information than is required for the purposes of this study. In fact, the highlighted elements could equally well be obtained by using the MOR tagger in the CHILDES CLAN tool, identifying nb=sg (singular), per=3 (3rd person) and tns=pres (present tense), along with the fact

---

[9] MPRO was originally designed as a spell and grammar checker and could, for example, indicate to a user that a particular section of the Duden grammar mentioned in Chapter 3 should be consulted in order to correct an error.

that the language is German (lng=germ) and that the lexical unit is the verb *werden* (lu=werden).[10]

This particular tag of *wird* was extracted manually after the processing of the entire text by MPRO was generously carried out in a pilot analysis by Paul Schmidt. Not only is an understanding of the complex method of analysis which is used by this program unnecessary for the present study, only the first item in the list (ori=wird) is required for the preliminary stage, following which we can advance to the second, qualitative assessment of the corpus.

A criticism of error annotation is levelled by Polio (Polio 1997: 117) who states that the fact that some errors can be categorised on more than one level is often ignored. Even when a multi-level analysis is carried out, however, a large amount of detail does not necessarily lead to clarification, as can be seen in (1). In this chapter, we propose to categorise what we term 'passive intent', which in this context means that the researcher believes a writer has attempted to express a passive, which may or may not be the result of conscious thought concerning grammatical form. This is operationalised in Chapter 5.1.

(1)   Die Erklärung   für \<**MoArInGn**\>   **diese** Phänomen   ist
       The explanation   for   this   phenomenon is
       (Mo = morphology, Ar = article, In = inflection, Gn = gender)

(example from Lüdeling 2006)

Both retaining the original corpus and the use of computerised methods allow researchers to treat the data according to their own research design, but in the study described here a non-annotated approach will be used, as explained in this section. In the second stage analysis, a qualitative examination of the original learner language will be carried out, providing evidence which cannot be deduced from a purely quantitative numerical count. Granger suggests that identifying passives is difficult unless a text has been annotated (Granger 1997). Although this is no doubt as true of German as of the English which Granger examines (the German auxiliaries *sein* and *werden* are plurifunctional, as is the verb 'be', as explained in Chapter 2.2.2.2), correct pragmatic

---

[10] I am grateful to Florence Myles for pointing out additional functionality of the CHILDES system.

usage in German cannot be identified via automated tagging alone; the corpus will require a closer, manual examination in order to extract only passive items, for which an annotation tool is not necessary. At the end of this process, the recall problem (Ball 1994) will have been avoided and precision will simultaneously have been assured.[11]

Having assessed possible options for automatically tagging the corpus in some way, at this juncture we drew on Granger's (2004: 127-128) elaboration of the tasks which learner corpora can enable researchers to undertake. Granger itemises four functions which are particularly well-suited to automation: *count*, *sort*, *annotate* and *compare*.[12] While the WordSmith 4 concordancing tools (Scott 2004) appeared to fit the bill as far as *counting* and *sorting* items in the data were concerned, further investigation was felt to be necessary before deciding whether or not to apply an *annotation* system to the corpus and whether or not an automated *comparison* would be possible in terms of responding to the research questions.

In this corpus-based approach, we are attempting to test a hypothesis based on evidence retrieved from students' written production. Even if assessment and measurement of a learner's progress is traditionally carried out by means of identifying errors, and easy identification of errors would be an argument for error annotation of a corpus, we are not concerned with measuring the number of errors here and equating that number with a stage in progression. Rather, the intention is to identify the tokens of passive use and evaluate that number *comparatively* across the corpus. In other words, this is a positive analysis of passive intent as defined earlier in this section and operationalised in Chapter 5.1, whereas error analysis by definition often concentrates on negative values, as indicated in the preceding discussion of annotation tools. We are not primarily benchmarking the learners' achievement against an unidentified completeness but attempting to map the learners' production of a particular construction.

For the reasons elaborated in this section, therefore, it was decided that no annotation would be undertaken. In addition, as one of the recurring aphorisms which appears automatically at the start of every concordancing procedure in WordSmith 4 states,

---

[11] To paraphrase Ball (1994), recall is how much of what is relevant that the system has managed to retrieve, while precision refers to the amount of material retrieved that is useful (i.e. relevant to the process in hand), and in turn enables the identification of missing and/or superfluous items. In a relatively small corpus, recall and precision are not difficult to achieve.

[12] These four terms and their derivatives are henceforth italicised in order to demonstrate the use of a strategy during the analytical process.

'Computer tools don't do the thinking'. A combination of automation and human intervention is applied to the corpus in the present study. An automated quantitative analysis of a collection of untagged texts by a concordancing program is considered to be justified, since we do not need recourse to a multi-layered, *annotated* corpus. Future pathways based on a tagged version are open to development and it is acknowledged that software tools other than the concordancer could have been used in the study described here.

4.5   Refining and Applying a Methodological Framework

The relative values of quantitative and qualitative analyses are well known and arguments for or against one or the other have been advanced in both the SLA literature (e.g. Larsen-Freeman & Long 1991; Mackey & Gass 2012) and the social science field (e.g. Bryman 2008) where quantitative and qualitative methods, respectively, have usually been favoured. Taking the most basic view that quantitative studies look for quantification in data and that conversely qualitative studies look at the quality of that data, a mixed-method type of analysis seems to fit the intention of this study, which is to use a learner corpus approach to illustrate the production of the German passive. The activities of *count*, *sort* and *compare* as described by Granger (2004: 127-128) can be carried out on the data in a learner corpus. Granger's fourth activity a*nnotate* is not taken as an option here, as explained in the previous section.[13]

Having argued for an unannotated database and the use of a concordancer to *count* and *sort* the data, the general aim of this study is repeated for clarity at this juncture, together with the 4 specific research questions. Following this, the precise methodological steps which were taken in this project are described, beginning with the assembly of the current LINCS corpus and its three constituent parts. In subsequent sections, a step-by-step account of the methods used to extract the information required by the research questions is given. Answers to questions 1, 2 and 3 will be sought by means of *counting* and *sorting* data, with the results being presented in Chapters 5 and 6. A response to question 4 will be offered in Chapter 6 and Chapter 7 where the *comparative* analysis (a version of CIA, as described in Chapter 2.6.2) will be described. The general research aim is expressed as follows:

---

[13] The term *annotate* is henceforward no longer italicised in the text.

To investigate patterns in the usage of the German passive by a cohort of learners with L1 English at different stages in their learning process in a formal learning environment and subsequently to suggest possible reasons for these patterns.

These four research questions drive the study:

1. Which forms of the German passive are produced in written interlanguage by adult learners with L1 E?
2. Do these forms vary over time?
3. Which forms of the passive are evident in the written production of adult NSs of German?
4. What similarities or differences can be observed between the use of the passive by NSs and NNSs in the respective subcorpora?

4.5.1 Benefits and Limitations of Electronic Data Collection and Analysis

The intention in the early part of this chapter has been to indicate that while the contribution of electronic means of analysis to corpus-based studies is indisputable it nevertheless is the case that the results of a quantitative frequency *count* are enhanced by further, manual analysis.

Adolphs (2006) indicates the value not only of fast access to data in an electronic corpus but also the speed at which those data can be analysed. In addition, data which are collected electronically can be easily retained for further analysis both by the original investigator and by others either for verification purposes or in order to initiate an alternative focus using the same data. A further benefit of using electronic analysis is that it facilitates studies of all types, whether cross-sectional, longitudinal, case study or corpus, and data can be measured both quantitatively and, to a certain extent, qualitatively, depending on the research question posed. On the other hand, Adolphs admits to 'limitations to using this method in isolation' (2006: 11) and suggests that the robustness (and therefore the validity and replicability) of the collected data can be ensured by careful selection of data and analytical method. In terms of a study of second language acquisition, learner corpus evidence can provide the examples to prove or disprove a hypothesis already generated. This evidence can also, in a qualitative paradigm, form the basis for the generation of the hypothesis in the first place. It is only

recently that learner corpus studies have been linked to the domain of SLA (Granger 2004: 134-135). This might be, for example, because one cannot explain learner language behaviour with examples from a corpus alone; the addition of a reflective learner response such as a think-aloud protocol might provide more clarity in this respect (e.g. Schauer & Adolphs 2006, who combine such a protocol with a corpus approach).

The benefits of computer-based analysis are no longer a subject for discussion, because technological advances continue to refine analytical methods, including making it possible to add some qualitative codes electronically rather than relying on manually coding texts such as the system suggested by Ellis & Barkhuizen (2005). Barlow (2005: 335) notes that '[t]he emphasis on frequency [...] means that software tools are an essential part of learner corpus research'. However, while an electronic frequency *count* forms the first stage of the present study, a manual *sort* of the initial *count* is used in order to disambiguate[14] passive auxiliaries from other forms of *werden* and *sein* before proceeding to the qualitative analysis. Working with a corpus of small size, this choice is justified. A detailed description of the process appears in Section 4.8.2.

In the first part of this chapter, it was shown that electronic systems offer improvements in terms of time-saving and accessibility for the collection and storage of learner language, and that a large amount of annotated data can be derived from a corpus by an automated tagging system but that the quantity should be measured with care against the research focus. This type of system is by nature objective, except in so far as the creator of each system has made some subjective decisions relating to the tagging of specific items. Advances in technology now enable the researcher to manipulate the coding system to suit his or her research purpose, as is the case with CHILDES, for example. The work of some researchers (e.g. Weinberger 2002), using an individually-designed tagging system devised to fulfil a specific research aim, initially remained in private hands perhaps because no suitable automated system existed at the time the study was undertaken.[15]

---

[14] Manual disambiguation is similarly described by Aijmer (2003).
[15] In Weinberger's case the tagging system was not initially automated; the tags were applied manually and checked by a second rater. It has been indicated in section 4.3 that Weinberger's corpus is now available via the FALKO platform.

Qualitative, descriptive methods, on the other hand, lead to results expressed subjectively and offer a richer perspective. In an attempt to make use of existing tools so that the results may be reliably interpreted, the data in the LINCS corpus are submitted to a generally available concordance program, WordSmith 4 (Scott 2004), in the quantitative analysis, followed by a qualitative analysis which is partially descriptive and partially based on a version of a formula which is in the public domain. This analysis will be described in Chapter 6.2.

The aim in using such methods is to address the criticism levelled against some SLA research by Ortega who points to the lack of uniformity in the measures used to define and categorise students' proficiency levels in studies of second language development (Ortega 2003). In the case of learner corpus studies, this may be attributed to the fact that learner corpora are still relatively young as a method of collecting data and a number of different research strategies are being applied. Ortega warns against 'employing statistical significance as the sole source of evidence for the presence or absence of a relationship' between items which are the objects of measurement (2003: 495) and this warning would seem to be useful in the case of the present study since the proficiency level of the majority of the subjects considered by Ortega appears to be similar to those whose work forms the bulk of the LINCS A corpus.[16] In fact, 62% of the cases reviewed by Ortega concerned the language of students of ESL and only 38% analysed the language of students learning a foreign language in the US. This lends further support to the point that was made earlier in this chapter that work on the acquisition of L2 English has been the focus of more studies than that of learners in a foreign-language learning situation, and lends more credence to the argument for carrying out this study which focuses on L2 acquisition in an instructed-SLA context in a language other than English. It was pointed out in Chapter 2 that FL and instructed-SLA contexts are considered by Ortega to be different in that FL learning often takes place over a longer time and the learners begin at a lower level (Ortega 2003: 512). Such a context variable may well produce a less valid result. Indeed, the systematic and standardised collection and evaluation of data according to principles outlined by the researcher are prerequisites for a representative, reliable and replicable study, and we aim to indicate the precision of these principles with a description of the members of the A cohort together with the context of the writing tasks they performed in section 4.6.

---

[16] In Ortega's study, the subjects were US college-level learners, approximately at the same educational level as UK post-secondary school learners at a university.

Ortega (2003) notes the lack of universal procedures for measuring interlanguage variation in the 25 studies on which she based her synthesis, pointing out that consistency is needed in order to draw reliable comparisons between the conclusions reached in the various studies. The relationship being examined here in this study is the one hypothesised to exist between the syntactically different ways of indicating the passive in German and the order in which their semantic equivalents emerge in the written language of adult learners. A statistical analysis in the form of a token *count* is used first and foremost as a means of quantitatively assessing the number of incidences of different types of German passive (the type/token relationship is established). The results of this test then allow us to *compare* and contrast the findings in a qualitative way in Chapter 6 and Chapter 7. The intention is not to investigate the development of the quality of written language by using a particular measurement tool such as measuring the mean length of sentences or mean length of clause. This type of measure is more suited to a study with a larger sample size and therefore larger corpus size than the one presented here. In any case, as Ortega points out, improved quality is not necessarily to be equated with increased sentence complexity e.g. a larger number of complex structures such as the passive. The concern here is the identification of syntactic and semantic development rather than with the evaluation of discourse quality.

The fact that Ortega discovers a marked increase in complexity when measuring the acquisition of language over time is, however, useful in this study of a complex construction since it suggests that the L1 E learners may produce not only periphrastic German passives but perhaps also other variations of the passive as time progresses. Evidence will be sought in the response to research question 4 in Chapter 7.

4.5.2   A Mixed-Methods Approach

In this section, the arguments upon which the choice of a mixed-methods analysis are based are presented.

According to Friedman (2012: 194), as long as researchers 'are honest (both to readers and themselves) regarding their value systems, possible biases and stances', the results of their investigations may be considered reliable. This would seem to support a robust, mixed-methods analysis. Ellis & Barkhuizen's appraisal of grounded theory explains a data-before-theory approach as follows:

> [I]n this *grounded theory* approach, the researcher begins with the data, and through its analysis (…) arrives at an understanding of the phenomenon under investigation. These themes and patterns do not simply jump out at the researcher – discovering them requires a systematic approach to analysis based on familiarity with related literature and research experience (Ellis & Barkhuizen 2005: 257).

It could be argued that the requirement for a researcher to be familiar with related literature suggests that at least some prior knowledge of theory is necessary before data collection begins. The research questions to be answered in this study emerged from theoretical considerations concerning developmental sequences in learner language found in the SLA literature, and it was determined that data collected cross-sectionally and longitudinally would make it possible to provide a response to those questions as a result of the quantitative analysis and to contribute to the understanding of language development by suggesting reasons for the patterns which are identified. In addition, a corpus-based approach permits further questions to be answered, should they emerge from the assembled data. Seen against this background, the decision which has been taken in the course of this study to undertake a type of qualitative *comparative* analysis[17] which follows a consideration of a quantitative *count* is justified.

Frequency *counts* as a means of assessing data are now considered.

4.5.2.1 Frequency Counts

It can safely be assumed that a quantitative *count* alone will not be appropriate for this study. Even researchers investigating frequency patterns do not propose 'a completely automated approach' (Rayson & Garside 2000: 3) without the support of a subsequent qualitative analysis.

Nevertheless, concordancers can function effectively as a primary means of quantifying data, in the present case to produce a numerical tally of passive tokens of different types, and can be supplemented by other, qualitative methods of interpretation.

---

[17] *Comparative* in the sense described in Chapter 2.6.2: *comparing* both the IL of the same group of learners cross-sectionally in a number of written exercises and also NS language with NNS language, both of them produced in an academic learning context.

Statistical methods such as chi-square tests which are often used in similar research paradigms are not suitable for use on small corpora; Larsen-Hall points out that a value of more than zero is required in every cell (Larsen-Hall 2012), and as can be seen from the tables in Chapter 5, zero scores were registered in several instances in the LINCS corpus data. We therefore argue that a combination of quantitative and qualitative analysis will suit the purpose of this study. In introducing qualitative analysis as a second stage, equal status is given to both types of analysis, neither 'avoid[ing] [n]or downplay[ing] statistical techniques' (Silverman 2000: 1) for those who believe quantitative research to be the 'gold standard' (Silverman 2000: 3), nor questioning the reliability and validity of qualitative research.

4.5.2.2 Cross-sectional or Longitudinal Analysis?

Cross-sectional studies facilitate *comparison* not only of the performance of subjects at one given point in time, simultaneously and on identical tasks (or as near-identical as possible, subject to certain constraints), but also at several points in chronological time. A cross-sectional case study approach is also beneficial in the handling of a relatively small corpus as an initial step towards testing a hypothesis. As Stake (2000: 422) says, 'How we learn from the singular case is related to how the case is like and unlike other cases (i.e., comparisons)' while longitudinal studies 'typically involve[s] observing the development of linguistic performance' (Larsen-Freeman & Long 1991: 11). Since one of the aims of this project is to observe the development of linguistic performance of the same individual over time, a longitudinal study based on the findings of a cross-sectional study will provide a suitable research paradigm. The trends in terms of individual performance might then be extrapolated to produce a hypothesis of the sequence of acquisition of the various ways of expressing the passive in German in this particular type of formal learning environment, although it is recognised that data from a relatively small corpus cannot be considered to have universal applicability. A larger data collection would have to be examined, but the results of the present study will, it is hoped, provide the basis on which future research may focus. If the same subjects can be tracked in both a cross-sectional and a longitudinal analytical exercise, the value of the analysis will be increased, since it will not only be possible to notice variation within an individual learner, but also to compare the performance in the interlanguage of several learners (the CIA described in Chapter 2.6.2) who should, according to their academic qualifications and training, be at the same general level of proficiency. Thus

the corpus data can be examined from a number of different perspectives, trends can be identified and further questions can be asked, as suggested in Section 4.4.2. Do the findings of the analysis have any implications for language pedagogy, for example?

In summary, as Larsen-Freeman & Long point out, 'the longitudinal or cross-sectional approach should not be associated exclusively with either [the quantitative or the qualitative – my addition] paradigm' (1991: 14). A combination of quantitative with qualitative methods, of cross-sectional (including individual case studies) with longitudinal investigations would seem to offer a suitable treatment for the data in the study described here.

4.6   Corpus Design

The importance of representativeness and replicability of a data collection has already been mentioned in Chapter 2.6.3.2 as integral to a reliable research study. To these two criteria it is proposed now to add a third: retrievability.

While the relatively small size of this corpus may give rise to criticism, in its support we can say that it has been designed to be a representative sample of the level of writing skills in German which can be expected at certain stages of a British applied languages university degree, and in fact the entrance requirements of the top grade in the school-leaving certificate in the FL to be studied are typical of any undergraduate degree programme in Languages in the UK, not just of the respondents in the A corpus (see Table 4.2). Conclusions drawn from this study could therefore be tested on a larger data sample which might include the written production of learners with the same qualification level in any instructed-SLA context.

It has been said (Biber 1990) that if texts of a comparable length prove to contain similar numbers of particular linguistic features, then they should be considered to be suitable items within a corpus. In the LINCS corpus, however, it is precisely the fact that this similarity may be absent which is interesting. Having taken account of as many variables as possible in order to achieve a reliable, systematic and therefore robust collection of data, it is the variation in the production of the passive which is under investigation here, whether it be overuse or underuse and the extent of its comparability to NS use. Nor is the breadth of range of the texts of concern here, but the number of

tokens of given types which each text exhibits. An electronic learner corpus enables us to retrieve relevant data.

After collating numerical quantification of tokens in the entire corpus, the data were examined qualitatively in order to seek evidence of patterns of production. In accordance with Myles' (2005) recommendation, the data were subjected to cross-sectional and longitudinal analysis. As a starting point, however, the first part of the study constituted a *comparative* case study of the written German of one L1 English learner (A8) and one L1 German learner (A7) on two specific exercises. The results of a case study approach were then examined together with those from a larger cross-sectional study of the same cohort of subjects to investigate whether the findings were replicated.

4.6.1 The Composition of the LINCS Corpus

The LINCS corpus comprises three sections, all containing written data, and currently amounts to approximately 84,180 words. The A corpus contains exercises collected from a cohort of university-level students of German, all of them except one with L1 E. Their written production was collected both synchronically and diachronically and amounts to approximately 46,900 words. The lengths of the exercises reflect the academic level of the writers, becoming exponentially longer over four years. Using the year of study as one of the variables in the production of texts could be said to contribute additional non-finite variability to a study, but in this case it is felt to be justified because the age of the participants is known, as are their university entrance qualifications and therefore their experience as students of languages (see Table 4.3).

The environment in which the texts were produced differs in that two were written with time constraints and two without. The Year 1 short exercise is typical of a first-semester essay written outside class time but with only two weeks' notice, the two second year examination questions reflect language produced under time constraints and approximately 18 months after the Year 1 exercise, and the Year 4 dissertation is a result of an individual's field work and research carried out during a six-month period abroad, produced first of all as a draft and finally submitted following comments on content but not on language by the student's supervisor. The final submitted version

provides the data sample for this study. Such variety is to be expected in a learner corpus as defined in Chapter 2.6.3.1.[18]

The B corpus contains only synchronic data totalling 34,121 words. The third element of the LINCS corpus, the C corpus, is a small subcorpus of input data consisting of 3,158 words, which is described in section 4.6.1.3.

4.6.1.1 The A Corpus

As far as standardisation of participants is concerned, all the students whose work is included in the A corpus were following the same university level course and had attended the same language and applied language classes for the duration of their university careers. They were taught by both NSs of German and NSs of English with pedagogical experience of grammar teaching to advanced or adult learners. Table 4.3 includes the entire A cohort, A7 being the only student with L1 G as opposed to L1 E. Table 4.3 identifies the qualifications and levels of attainment in German of the students in the A cohort when they commenced their university education. There are differences between the admission criteria in England and in Scotland, and since the A cohort comprises students from both countries as well as one NS from Germany, some additional information is included below the table in order to interpret it fully.

---

[18] For clarity, Granger's definition is repeated here. Learner corpora are 'electronic collections of spoken or written texts produced by foreign or second language learners in a variety of language settings' (Granger 2002: 124).

Student id	Gender	Age at 1.9.**[19]	Grade at GCSE/Standard (German)	Grade at H/AS	Grade at AH/AL
A1	F	18.0	1	A	A
A2	F	18.09	A*	A	A
A3	F	18.04	1	B	C
A4	F	18.05	A*	N/A	B
A5	M	18.0	A*	A	A
A6	M	18.05	1	A	A
A7	F	19.09			13/20
A8	F	18.02	A*	A	A
A9	F	17.11	1	A	A

Table 4.3: A cohort prior to admission to university

N/A indicates no qualification at this intermediate level between GCSE and A level.

1.   In England, AS level examinations are taken at the age of 17, an age corresponding to the age at which school students in Scotland may sit Higher (H) examinations. Pupils in Scotland may then opt to go on to sit Advanced Highers (AH) at the same age as pupils in England sit A levels (AL).

2.   'Age' is shown as it is indicated by the applicant on the Universities and Colleges Admissions Services (UCAS) application form. 'Grades' are on a scale of A-C or 1-6 where A or 1 is the highest (A* being the highest award at 16+ in England).

3.   School pupils entering higher education in the UK who do not take a year out at the end of their schooling are usually 18 years of age. A9, the youngest in the cohort, reached her 18th birthday shortly after commencing her university course. A4 entered the university a year earlier than the rest of the cohort but suspended studies for 12 months between years 1 and 2. The inconsistency in the age of A7 is accounted for by the fact that education in Germany commences a year later than in the UK and therefore terminates at the age of 19 rather than 18. In terms of comparability, she has been in education for the same length of time as the rest of the A cohort.

---

[19] ** Double stars are inserted in the date format to protect the anonymity of the learners.

4.   With two exceptions, the L1 E students commenced their university education with the top grade in German.

5.   A7's grade mark of 13/20, is the *Fachhochschulreife* (the school-leaving certificate) awarded by Land Nordrhein-Westfalen. This mark roughly equates to a mid-B grade in the British system. In addition, this student was awarded 12/20 for English. While 12/20 does not equate to a British Grade A, and is therefore a lower level than the FL competence of the L1 E cohort, it is not her FL competence which is useful in this study but her identity as a NS of German who is temporarily not using her mother tongue in day-to-day situations. A7 also achieved an IELTS score of 8.[20] It is assumed that learners with these university-entry-level qualifications will have already received a certain amount of formal instruction of the two basic syntactic forms of the passive, since these figure in the list of prerequisites from various examination boards.

4.6.1.2 The B Corpus

There are no such detailed records of the members of the B cohort as are shown in Table 4.3 for the A cohort. The B corpus may be described as an L1 corpus i.e. the users all have the same mother tongue, as opposed to a learner corpus produced by learners with different mother tongues all learning the same L2, who are often the subjects in learner corpora of English as described in section 2.6.3. The B corpus is not a learner corpus of the same type as the A corpus since the students in the B cohort are classified rather as users than learners, but it is a parallel corpus introduced for the purpose of carrying out a type of CA. The term CIA was introduced in Chapter 2.6.2, and is useful in the attempt to avoid the Comparative Fallacy (Bley-Vroman 1983), that is, ensuring that a learner's interlanguage is not judged according to NS norms but as an entity in its own right.

The B corpus comprises just one written text from each respondent; there are no longitudinal data. The texts were offered to the LINCS database under the guidance of the academic moderator of the doctoral workshop at a European Second Language Association (EUROSLA) conference at which earlier stages of this research were

---

[20] International English Language Testing System (IELTS) has 9 rating bands. Band 8 classifies a '**Very good user:** has fully operational command of the language with only occasional unsystematic inaccuracies and inappropriacies' (British Council 2013).

presented. The L1 G users were all university students and their written production was considered to be at an academic level and written in a similar academic genre to that of the A cohort in year 4.

The B corpus not only includes data from an external group of L1 G users, but also, after the initial preliminary study, the production of A7 from the A cohort. This respondent's production is assessed in comparison to both the A and the B cohorts.

4.6.1.3 The C Corpus

The small C corpus comprises 12 texts of varying lengths which were supplied by the tutor during the second year of study of the A cohort, that is respondents A1-A9, including the one L1 G (A7) who belonged to this cohort of students. The texts are all taken from the German press or internet, with the exception of one tutor-generated solution, and were used to generate student responses in the form of written summaries, essays or cloze test exercises. These texts were supplemented by formal exercises taken from Dreyer & Schmitt's grammar (Dreyer & Schmitt 2001). Neither the input nor the students' output in these grammar exercises constitutes part of the corpus because they represent a specific type of form-focused drill which is not included in the present study.

Adult learners such as those in the A cohort undertake a certain amount of naturalistic exposure to the FL during a six-month period spent in a German-speaking country, and during this time no controls on input are possible. The unknown quantity in the study is the quality and quantity of German input during this time. The effects of this naturalistic input on the learners' holistic proficiency are not measured in this project although they are assumed to have influenced the forms of the passive produced in the students' dissertations. The chief aim of this study is to observe and comment on any changes in the A cohort's production of the passive over a period of time; this will include both the passives produced by the L1 E learners and A7, the single L1 G user in the A cohort.

A causal-correlational study comparing input of the passive with users' output is not undertaken here except in so far as limited conclusions can be drawn regarding the frequency of the two syntactic passive types which are *counted* in the input in year 2.

4.7 The Corpus Format

The LINCS corpus is stored in Word-readable document format. Full concordance files are stored separately, firstly because they cannot be displayed in isolation from the WordSmith 4 program (Scott 2004) which created them and which is available for purchase. Secondly, the original format is retained until a decision has been made regarding the collection to which the corpus will be submitted. The exception to the Word format is A7's Year 4 exercise which appears as a .txt document, this being the format in which it was originally presented. Screenshot examples of the relevant processes involved in the concordancing procedure are given in section 4.8.2.1.

4.7.1 Preparation for Analysis

We devised a permission form in 2008, to be signed by students at the beginning of the academic year, indicating their agreement to their performance (written or spoken) being used for research purposes (Appendix C). Prior to 2008, students were assumed to have agreed to this by having sent the electronic files to the author. If a subject did not respond to the request to send an electronic file containing the Year 4 dissertation, their Year 1 and 2 work, where it had been possible to retrieve it, was not included in the corpus. In the case of student A4, although there was no Year 1 exercise, the Year 4 dissertation was contributed to the data collection and the Year 2 examination responses were therefore added to the corpus as well. Electronic analysis, by definition, necessitates the data being available in machine-readable format before being entered into the processing program. As a first step, therefore, the handwritten texts had to be word processed by the researcher in order to reproduce them as computerised files.

4.7.1.1 Text Editing

Only the Year 4 dissertations were available as electronic documents. In preparing these for concordancing, only continuous text was retained, together with footnotes and endnotes, as they were deemed to have been composed by the writer and therefore to exemplify his or her own language production. Bibliographies and reference lists, where appended to the original documents, were omitted, as were Tables of Contents. Appendices were retained if they comprised continuous text or questionnaires devised by the author, but were deleted if in tabulated form.

Electronic copies of the handwritten Year 1 short essays and the Year 2 examination answers were produced, ensuring that the original orthography was retained. No corrections were made. The titles of the exercises in year 2 were omitted because they were given in the examination rubric. The texts in the C corpus were grouped according to topic for ease of recognition and retrieval in future. The topics equate to the pedagogical themes addressed in the student's learning programme: Government Systems (3 texts), Education (2 texts), Social Welfare (5 texts), Media (1 text), and a single Class Test solution. Titles were included but references to the source of each text were not. This information was retained in a separate file for future reference. The resulting corpus was saved as a .txt file in the same format as the corpora of students' work, in preparation for analysis by the concordancing program.

4.7.1.2 Anonymising

To comply with data protection requirements, the electronic documents were manually anonymised. All references to the identity and affiliation of the subjects were removed by carrying out an electronic search and delete operation. Each item was then copied to produce a working document bearing ownership of the project author only. Further editing will probably be necessary depending on the norms of the data collection which ultimately accepts the LINCS corpus.

Any instances of the passive which were extracted during the concordancing process but were identified as quotations taken from other sources were excluded from the analysis; this applied specifically to the Year 4 dissertations and the texts submitted by the members of the B cohort of L1 G users. Arguments supported by quotations are expected in this type of extended dissertation writing, whereas they were not required in either the Year 1 exercise or the timed examination answers.

4.7.1.3 File Conversion

WordSmith 4 (Scott 2004) requires that files are in machine-readable .txt format. The final act of preparation thus entailed saving the data as individual text files with the .txt suffix. The LINCS corpus comprises the collection of these files in three subsections.

4.8 The Process of Analysis

The three analytical processes, the case study, the quantitative frequency *count* and the qualitative analysis are described in this section.

4.8.1 The Case Study

Studying one or more cases on an individual basis enables a clear focus on specifics as opposed to the identification of a more general trend which may result from a longitudinal or cross-sectional study. A case study approach is therefore beneficial in the handling of a relatively small corpus as an initial step towards verifying a hypothesis in the same way as longitudinal and cross-sectional studies are preceded by in-depth investigations of one or two learners, in order to identify issues which warrant further examination. It must be remembered that case study results cannot be generalised to other cases particularly if the sample size is small. 'The purpose of a case report is not to represent the world, but to represent the case' (Stake 2000: 448). Observations are made here regarding differences in the production of passives in the IL of a particular learner cohort. Further research on a larger sample is necessary before significant claims can be made.[21]

Although Duff remarks that case study research is no longer so focused on the 'technical aspects of learners' language' (Duff 2012: 100) and specifically mentions syntactic development among these technical aspects, we contend that a case study provides a strategically valuable tool in a mixed-methods analysis such as is used here. Beginning with a *comparative* study, the results of that study were generalised towards a new theoretical position by testing the hypothesis advanced in Section 2.4 against a wider empirical sample. This hypothesis suggests that the *sein* passive is acquired after the *werden* passive, even though there is evidence in the data that the *sein* passive may be overused in the initial developmental stages. The results of an initial *comparison* of an exercise produced by a student with L1 E with a similar exercise produced by a student with L1 G were the precursors of a study of a more numerous set of respondents in order to assess similarities and differences in the written production of two different language populations, in response to research questions 1, 2 and 3, which could then provide support for a response to question 4.

---

[21] See also Section 4.5.2.2.

The initial *comparison* of the written production of two students, one with L1 E and the other with L1 G, who were following a university course in Languages at a British university, aims to investigate the possibility, observed during a preliminary examination of the data, of over-representation of the passive in German by L1 E learners at an intermediate stage of production which subsequently realigns more closely with native speaker usage at a more advanced level. The student with L1 G (A7) in the first case study exercise is taken to represent a typical NS at this stage in her academic career. She had been educated in the German school system, entering the British tertiary education system with the same mother tongue qualifications as would have given her entry to the German higher education system and with the addition of a qualification in English (her FL) equating to the FL qualification of the L1 E learners in German (their FL). The student with L1 E, A8, achieved a grade A at 'A' level in German and A7 an IELTS score of 8, as mentioned in Section 4.6.1.1, suggesting that their externally assessed competence in the FL was roughly equivalent and that they had met the foreign language admissions requirements for their course. As will be seen in Chapter 5, following the preliminary study student A7 is included in the L1 G subcorpus as an example of a NS as well as in a comparison with the rest of her student cohort of L1 E learners (students A1-A6 and A8-A9) in a contrastive interlanguage analysis. Both students A7 and A8 were considered well qualified for a languages course in interpreting and translating where mother tongue proficiency and foreign language fluency carry equal value rating. The difference in age of one year between the two participants is explained by the age at which schooling is commenced in Germany, typically one year later than in Britain.

4.8.2   The Quantitative Analysis

In this section, the process of concordancing adopted in this study is described. The beneficial or negative aspects of the process are alluded to. The creation of a type/token relationship was the goal, where types of passive (the various grammatical forms of the auxiliaries *werden* and *sein*) were subdivided into the total number of tokens of each of those types. If we agree with Richards (1987: 203), who suggests that 'TTRs (type-token ratios; my clarification) calculated from a large number of tokens will generally be lower than those calculated from a smaller number', we might expect that the number of passive tokens will decrease over time in a learner's production. Richards is suggesting that as a child's vocabulary increases, so the tendency for repetition

decreases. It may be the case that, in a similar way, learners use the passive less frequently over time if they acquire alternative means of expressing it. However, the quantitative *count* of the type/token relationship which is the goal of the initial concordance will not address such a qualitative question; that discussion appears in Chapter 7.

4.8.2.1 The Concordance Process

The concordancer was set at -5 to +5, indicating that a maximum of five words to the right and to the left of the search term would be shown on the concordance screen. While this parameter would usually be sufficient in English to identify both syntactic and semantic information from the concordance, the more complex sentence structure of German will not always allow such easy identification. It was noted in Chapter 3 that the German periphrastic passive is constructed using AUX and PP, but it is not possible to estimate how many lexical items will occur between the two in a standard V2 sentence. Compare, for example, the following sentences with predicates of differing lengths from student B1, one of the respondents in the B cohort:

(2a) In der Süddeutschen Zeitung **wird** Bundeskanzlerin Angela Merkel in der Ausgabe vom 19. August 2008 mit der Forderung nach einer „Bildungsrepublik" **zitiert**.

(2b) In the Süddeutsche Zeitung **becomes** Chancellor Angela Merkel in the edition of 19th August 2008 with a demand for an "Education Republic" **cited**.

(2c) In the 19th August 2008 edition of the Süddeutsche Zeitung, Chancellor Angela Merkel **is cited** as demanding a republic with a strong focus on education.

(3a) Im Zuge der Bildungsexpansion **wurden** bestehende Schulen **ausgebaut.**

(3b) In the course of the expansion in education **became** existing schools **refurbished**.

(3c) Already-existing schools **were refurbished** as part of the expansion in education.

2(a) and 3(a) illustrate the fact that the auxiliary and the past participle in a German sentence are not systematically juxtaposed, whereas in the English equivalents 2(c) and

3(c) they are. Hence in the case of syntactically complex sentences such as 2(a) a manual qualitative assessment is required where there is no visible indication of passive intent, as defined in Section 4.4.4, in the resulting -5 to +5 concordance because the meaning of a German sentence containing two verbal elements is not fully revealed until the final element appears. In Figure 4.2, N2, there is similarly no clear intent to be determined from the concordance as it stands. The process of reviewing the full source text in this circumstance is described below, following Figure 4.2. The results of the second stage, qualitative analysis will be used to assess the students' production of passives by *comparing* it to that of NSs of German, and support for the reliability of the analyst's judgement will be provided with reference to a check carried out by an NS interrater.

The examples (2a – 2c) given for the comparison of sentence structure above may also be compared with the concordance in Figure 4.2, where a lack of clarity exists for N11 and N18 – completion of either of the two tokens of *sind* with PP or with an adjectival form would be feasible but the intent cannot be determined by the existing concordance. On the other hand in a standard V2 sentence such as N24, it is simple to detect a copula *sein*:

(4)	und	die meisten	davon	sind	osteuropäisch
	and	most	of them	are	East-European.

This particular concordance item would be readily deleted, as would the V-final N6, where the verb phrase *geworden sind*, occasioned in all probability by a preceding relative pronoun which is unseen in the +5/-5 concordance, is not a passive but a past tense.
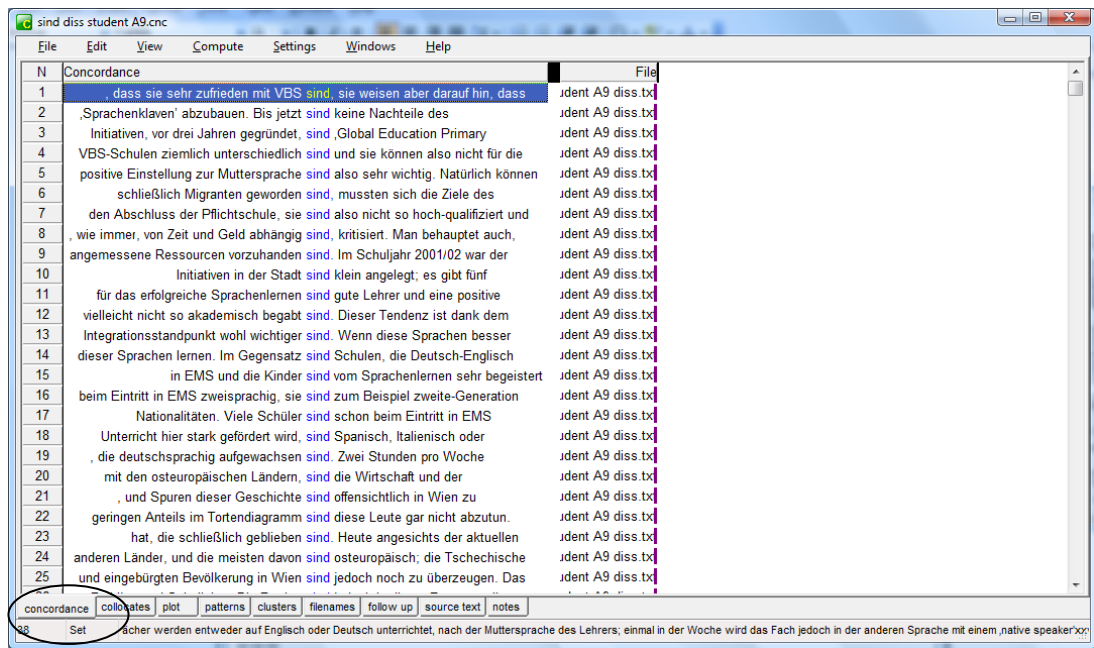
Figure 4.2: Concordance of *sind*, student A9

The WordSmith 4 tools (Scott 2004) were used to identify instances of the passive searching for the following verbal forms: *wird¸ werden, wurde, wurden, ist, sind, war, waren,* and *worden* with various tenses of *sein*. This search revealed not only periphrastic passives but also those forms of the construction involving use of *sein + zu + infinitive*. Subsequently the decision was made also to include the other alternatives to the passive which were mentioned in Chapter 3, the reflexive *sich lassen* (= to allow something to be done) and the indefinite pronoun *man* with a verb in the active voice. In the case of *sich lassen*, the forms *läßt/lässt sich, lassen sich, ließ/liess sich and liessen sich* were presented to the concordance program to take account of any misspellings in the light of the German spelling reforms. A numerical token *count* of each type was obtained. While the argument has been advanced, for example by Ellis & Barkhuizen (2005), that a straightforward quantitative word frequency *count* has serious limitations as a valid research metric, particularly because item frequency has often been measured against tokens which are not of a similar word class, in this case it is used purely as a first step towards identifying passives, as a precursor to a secondary, qualitative analysis. In the case of the German passive, where the auxiliaries used have more than one other grammatical function, a token *count* will not give a reliable indication of passive use: a further analysis will be necessary, as we have explained earlier in this chapter.

During the concordance process, the first concordance list which is produced contains all instances of the search word with the total numerical count shown bottom left of the screen (Figure 4.2). Each file was saved with the name of the search item followed by the suffix denoting the written exercise from which the concordance was taken plus the student id reference, e.g. *sind* diss student A9 (Figure 4.2). This was repeated eight times for each student respondent, the searchwords being the items in Table 5.5. In this program, once the initial cycle of repetition is completed, the searchwords are automatically retained in the drop-down menu, allowing the analyst simply to click on the next and subsequent items to create a new search and to enhance the fluidity of the exercise.

A manual review of all the items was undertaken, assessing for passive intent as defined above in Section 4.4.4. In cases where it was not possible to ascertain from the concord line alone whether a passive had been attempted or not, double clicking on the concordance line allows the entire reference section of the source text to be viewed. Close examination of the sentence which includes the highlighted search word (*sind*) reveals the syntactic intent of the searchword. Figure 4.3 shows the results of clicking through from the first item (N1) on the concordance list in Figure 4.2.
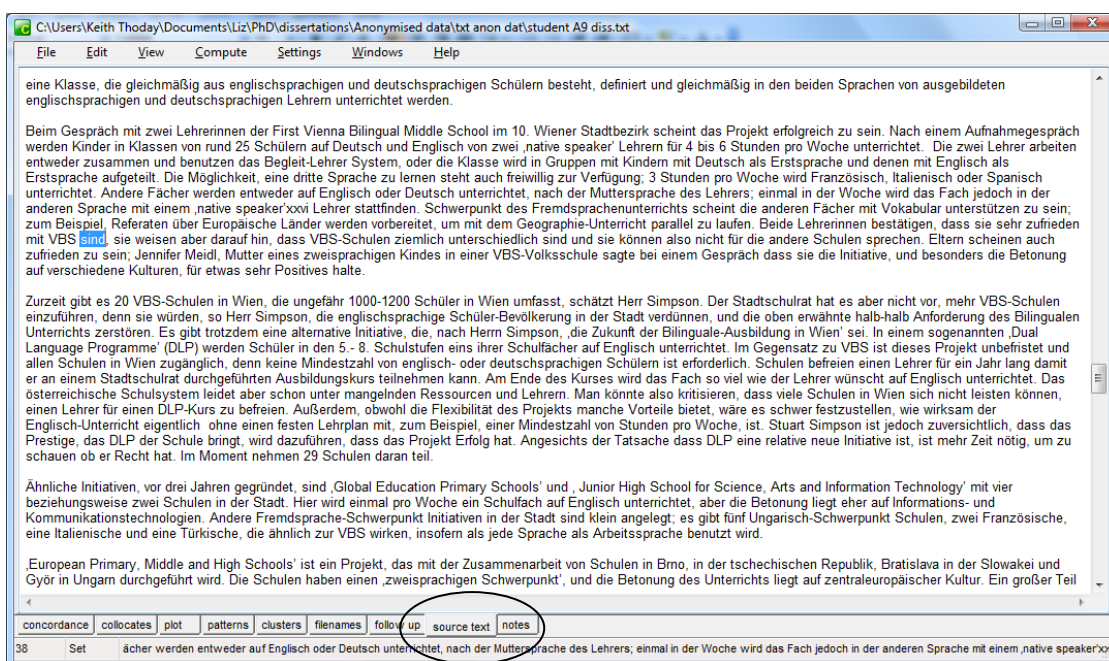
Figure 4.3: Source text, student A9, concordance *sind*

114

Those items which were not assessed as attempts at the passive were deleted.[22]   The files were saved in the same way, with the addition of PASS in the file name e.g. *sind diss PASS student A9*. The format of the resulting file (Figure 4.4) makes passive items clearly identifiable on subsequent visits to the data, for example during the qualitative analysis. In addition, to clarify the information available on the screen, only columns containing information relating to the basic token *count* were retained; extraneous columns to the right were hidden (indicated by the dotted purple line).[23]

Figure 4.4: *sind* **passive intent, student A9**

Retaining two files in this way will enable the original text (Figure 4.3) to be made available for continued research. A future project might, for example, focus on a *comparison* of copula and auxiliary use of *sein*, in which case a researcher might wish to use the original file as a starting point, retaining all instances of *sein* but deleting all passives. A new concordance file would then be produced using the original (shown in Figure 4.3) as a starting point. For the purposes of the present study, the information retained in Figure 4.4 ensures that the results can be verified in a fairly straightforward manner.

---

[22] The process of deleting is *Edit>Deleting>Delete.*
[23] Hidden columns such as 'position of word in sentence'.

The ability to view all concordance entries, even if greyed out, ensures a measure of controlling against errors in the editing process: in the event of a line being inadvertently deleted, it can be reinstated.[24]

A decision had to be taken whether to include in the frequency *count* any duplicated combinations of AUX + PP (i.e. repeated use of the same verb with passive syntax) as one type, or whether to *count* these repetitions as separate tokens. As Ellis & Barkhuizen (2005: 83) point out, *counting* types rather than tokens would reduce the overall item *count*. Since the corpus size in the present study is relatively limited, it was judged more appropriate to include every example, to avoid reducing the amount of available data still further. In Chapter 6.6, we shall mention a possible reason for learners using the same grammatical form more than once. The principle of *counting* every item listed in Table 5.5, even if duplicated, was applied to the entire set of calculations.

4.8.2.2 Measures of Accuracy

There is some debate as to how linguistic accuracy should be measured. What is required in this study is a measure of tokens which is independent of sample size. Hence the decision was taken following the initial case study to calculate the number of passive tokens per 500 words. This ratio was considered as a suitable metric for application to texts of varying lengths in which the passive, according to expectations considered in Chapter 3, would not appear frequently. It must be borne in mind, however, that a decline in token frequency may be observed in more advanced learner language if the learner produces alternative constructions which emerge as a result of his or her developing L2 grammar. This point will be addressed again in Chapter 7.

In this type of analysis, accuracy measures such as those discussed by Polio (1997) need not be considered. These measures include looking for error-free T-units[25] or devising a scaled error classification for passive use. The intention at this point in the quantitative analysis is neither to evaluate the gravity of errors nor to provide a subjective assessment of writing performance but to categorise attempts at the passive using an

---

[24] *Edit>Restore deleted>Restore* allows the line to be reinstated.
[25] T-unit: a linguistic term for the shortest sentence or independent clause which can stand on its own (Hunt 1965).

objective type/token ratio. The qualitative assessment of the semantic difference between the *werden* and *sein* passive alternations is carried out in the longitudinal analysis. An attempt is made to avoid subjectively assessed rater preference by offering a number of passive instances, selected at random from the corpora, to be rated independently as correct/incorrect (see Appendices D and E). A NS of German and a NS of English were asked in a questionnaire to confirm the analyst's L1 E assessment of the correct grammatical form of some of the identified tokens.

Polio (1997) attempted to ascribe a level of reliability to various types of assessment used in ESL essays, having found this unavailable in previous studies. The present study, on the other hand, is concerned with the correct or incorrect interpretation of passive intent, as described earlier in this chapter, to which no scoring system has been applied. The objective decision of the rater is based on the researcher's experience of teaching German as a foreign language. There is little likelihood of ambiguity in defining the correctness of passive syntax since the verb morphology is prescribed, as explained in Chapter 3.3.1. Semantic use of the passive, on the other hand, will be one of the subjects under scrutiny in the longitudinal analysis.

4.8.3   The Qualitative Analysis

In the introduction to Chapter 4, the arguments for supplementing the quantitative analysis with a qualitative assessment were advanced. Full details of the qualitative analysis are given in Chapter 6 but a brief introduction follows here.

4.8.3.1 Target-like Choice Analysis

The target-like choice analysis which was carried out following the quantitative cross-sectional analysis assessed the quality in terms of the percentage of correct choices of AUX (*werden* or *sein*) produced by the L1 E cohort and the B respondents. The learners' production was then *compared* using a type of Contrastive Interlanguage Analysis as described in Chapter 2.6.2. The correctness of morphological form of either AUX or PP was not considered at this stage but addressed in the *comparison* of NNS passives and their NS equivalents which is described in Chapter 7. Assessing correctness implies that errors were identified and this is certainly the case. However, as was explained in Section 4.4.1, the errors are not the primary focus here. Identifying

them simply allows us to refine the area of study still further and to focus on the positive choices which the learners have made.

The mention of 'choice' suggests that learners consider options when faced with using a passive in German. This may be the case in untimed exercises where metalinguistic knowledge[26] can be drawn upon, but if for example that metalinguistic knowledge is lacking in respect of the *sein* passive, then a learner clearly cannot exploit it. Equally, a learner may only focus on producing a grammatically correct form of a construction if the exercise he or she is undertaking specifically encourages such focus, such as in the second examination question in the data sample which contains a passive in the title. Another possibility is that learners may exhibit avoidance strategies and choose to express a passive in a different way, particularly if they regard it as a difficult construction. The discussion in Chapter 7 will return to these points in connection with the investigation of passive usage by NNSs and NSs of German.

When comparing the token *counts* and subsequently what we have termed the target-like choice analysis in the A corpus and the B corpus, attempts which exhibit use of a modal verb were not included in the analysis, which focuses on periphrastic passive use, even though, as was indicated in Chapter 3.4, this passive type requires no morphological change in PP and it might therefore be expected to be produced more readily in a learner's IL. Passive expressions using *sein* + *zu* + INFIN, on the other hand, were included in the analysis, although they differ from periphrastic passives in that *sein* is used and *werden* is intended. Duden (1998) treats *sein* + *zu* + INFIN as an alternative form to a *werden* passive. This type will be mentioned again in Chapter 7 not least because such expressions appear in the production of the B cohort and examples of them are therefore useful in responding to research questions 3 and 4.

4.8.3.2 Additional Analytical Treatments

The students in the A cohort responded to a questionnaire[27] on their comprehension and production of the German passive and also to a grammaticality judgement exercise

---

[26] Krashen (1982: 90) indicates that it is during timed exercises that learners are able to monitor their grammatical accuracy, but notes that a learner must be 'thinking about correctness or focussed on form' i.e. conscious thought must be employed; merely having time available will not be enough.

[27] See Appendix A.

(GJE).[28]   These two items are not discussed in this thesis although a preliminary analysis of the results is mentioned in Chapter 7.

4.9   Review of Chapter 4

Chapter 4, which is principally concerned with determining the analytical methods which are applied in this study, commenced with an overview of learner corpora in German, in support of using a learner corpus approach here to offer a contribution to an area of study which is as yet under-represented.   The choice of a concordancing program as a suitable analytical tool for this study was justified following an assessment of a number of available tagging systems.   The value of both quantitative and qualitative analysis was discussed and arguments for a mixed-method analytical framework were advanced.

The chapter continued with a precise description of the subjects who contributed to the LINCS corpus and the three subcorpora which make up the corpus were presented, together with an assessment of the reliability of the data in terms of their variables. After describing the method of assembling the data, the first of the three investigations, the case study, was introduced.   Following the presentation of the case study, the process of concordancing was explained, illustrated by examples from the data.   The benefits of electronic access to the data in terms of speed, ease and flexibility in the preparation and retrieval of the data for the quantitative frequency *count* via the concordancing process are self-evident and the choice of an electronic measuring tool would thus seem to be justified.   Thirdly, the qualitative methodology which was applied to the L1 E and the L1 G corpora was introduced.   A form of CIA as described in Chapter 2.6.2 was proposed.

The first stage of the analytical process, the quantitative studies of the data in all three subcorpora, will now be described in Chapter 5.   Analysing the LINCS learner corpus will provide some evidence of the patterns of use of the *werden* passive and the *sein* passive in the production of the learners in this particular formal learning environment, as suggested in Chapter 2.4.   As far as it has been ascertained, an investigation of this type has not so far been carried out.

---

[28] See Appendix B.

CHAPTER 5 - QUANTITATIVE DATA ANALYSIS: FREQUENCY COUNTS OF THE PASSIVE

5.1   Introduction

The methodological arguments and justifications for selecting the types of data to be analysed in this study have been stated in Chapter 4. In keeping with Ellis' statement that 'the type of data used closely reflects the research goals' (Ellis 1994: 671), this current chapter commences with a reminder of the research goals as they were set out in Chapter 1.7, and continues with an examination of the frequency of occurrence of the syntactic passive in the corpus. It should be pointed out that an in-depth discussion of the results of the quantitative *count* and *sort* process will not be carried out in Chapter 5. Instead, the quantitative results are presented and are accompanied by short comments drawing attention to particular aspects of a learner's production which are noteworthy in terms of the research questions posed. A more detailed assessment of the results of the qualitative analysis is the subject of Chapter 6 and Chapter 7, to which the results of the frequency *count* will contribute.

For reference purposes, the research questions which drive this analysis are repeated here, preceded by a reiteration of the general research aim, as follows:

To investigate patterns in the usage of the German passive by a cohort of learners with L1 English at different stages in their learning process in a formal learning environment and subsequently to suggest possible reasons for these patterns.

The reasons for attempting to identify patterns of development in the acquisition of this complex structure relate to the research on developmental sequences which was discussed in Chapter 2.5.3.3. The aim here is to extract examples of passive intent from the corpus, that is, to identify the instances when the writers attempted to produce a passive, and to evaluate their success in producing NS-like syntax. A cross-sectional analysis will demonstrate the extent to which each item is used by each respondent at a given point in time. By definition, if changes in usage are to be identified, there must be observable longitudinal change, hence the need for data also collected over time. An electronic collection of data such as the one on which this study is based offers the

opportunity to carry out searches relatively quickly and because the corpus contains the written production of the same cohort of students collected both synchronically and diachronically, both cross-sectional and longitudinal analyses are possible.

Subsequently, the researcher should be able to look for support for a pre-existing hypothesis as well as for an additional one which may emerge from studying the data. In this study, as was explained in Chapter 2.4, the primary hypothesis being tested is based on a review of L1 studies of the acquisition of the German passive and comparing them with evidence from the pedagogical practice of teaching German as L2. It relates to the existence, or not, of step-wise development in a learner's use of the construction.

Having established the overall research focus of the study, two research aims were established.

Research aim 1:

> To contribute to the theoretical discussion concerning the relationship between SLA and second language pedagogy.

Research aim 2:

> To construct an embryonic learner corpus and offer data to a wider research population.

While the creation of the corpus in accordance with research aim 2 was described in Chapter 4.6, the achievement of research aim 1 drives the analyses which are reported in Chapters 5, 6 and 7. Four questions were the starting point for the preliminary analysis, and for the subsequent phases of the study, namely:

1. Which forms of the German passive are produced in written interlanguage by adult learners with L1 E?
2. Do these forms vary over time?
3. Which forms of the passive are evident in the written production of adult NSs of German?

4. What similarities or differences can be observed between the use of the passive by NSs and NNSs in the respective subcorpora?

The answers to questions 1, 2 and 3 were sought in the first instance via a quantitative analysis of the production of two respondents in a case study and subsequently in a frequency *count* extracted from a larger data sample, by carrying out *count* and *sort* procedures. These procedures were then repeated on a data sample from a second population of respondents and finally the results were used to support a *comparative* analysis in response to question 4.

5.2   Preliminary Analysis

The preliminary analysis examines data from two respondents, one with L1 E (A8) and one with L1 G (A7), collected both cross-sectionally and longitudinally. This is followed by a quantitative analysis in four parts. Firstly, the entire A corpus, including A7 and A8 from the preliminary analysis, is examined from a cross-sectional perspective, carrying out a quantitative study of a more extensive collection of data and commenting on the results of specific individuals. Secondly, this is followed by a longitudinal analysis of the same dataset. Thirdly, a quantitative analysis of the B corpus of L1 G users is carried out. Some reflections on the production of A7 in relation to both the A and the B corpora are included. The chapter concludes with an assessment of the results of the *count* and *sort* procedures and looks forward to Chapters 6 and 7 where the statistics generated by this quantitative analysis form the basis of a qualitative study of the datasets of the L1 E learners and the L1 G users.

In Chapter 4, it was indicated that three (*count*, *sort*, *compare*) of Granger's four points denoting the processes which learner corpora allow us to carry out are well-suited to the aims of this project, that is, to identify production of the passive in second language learner written production.[1] *Counting* and *sorting* not only allow us subsequently to *compare* learner with learner cross-sectionally, these processes also form the building blocks for the construction of an hypothesised series of stages through which learners pass as they attempt to master the German passive. Additional information is collected by *comparing* one performance with another, longitudinally.

---

[1] The fourth element, '*annotate*' was discounted for the purposes of this study, as was reported in Chapter 4.5.

Two forms of the passive were chosen for the preliminary analysis, the 3pers sing pres of *werden* (*wird*) and the same form of *sein* (*ist*). The number of *wird* passives and *ist* passives was *counted* electronically by means of a concordancer. This initial *count* then enabled a manual *sort,* disambiguating the passive auxiliaries from the copula verbs. Although computerised programs which can identify syntactic structures are either available (e.g. using the CHILDES system) or under development in various research institutions (e.g. the group led by Lüdeling in Berlin), a manual *sort* to identify tokens of the passive from the retrieved concordance was chosen for the present study because of the relatively small size of the corpus, as discussed in Chapter 4. Following this *sort*, the numerical *count* of passives of these types in the production of one respondent with L1 E (A8) was *compared* to that of one respondent with L1 G (A7); both of them were studying in the same student cohort.

5.2.1   Quantitative Concordance Part 1

The data for the preliminary analysis comprised two items of written text in German, one produced in year 2 and the second produced in year 4, by each of two students, one with L1 E (A8) and one with L1 G (A7). The texts were selected from the written production of the A cohort after a preliminary, cursory observation of the students' second and fourth year exercises. Some obvious differences in terms of the number of incidences of the syntactic passive were observed, giving rise to the strong suggestion that usage of the passive by a learner of German differs from that of a NS user. The decision was taken to examine these four written items more closely as a first step, before extending the analysis to include the examination of the entire corpus. This step would enable, firstly, a comparison to be made between the production of A8 and A7, secondly, the next step in the investigation to be planned and, thirdly, the validity of the researcher's initial impression to be tested. A type/token representation was used, where *type* refers to the search term entered in the concordancer and *token* the number of instances of that type per unit of production, as explained in Chapter 4. Instances of the passive were calculated according to the number of passive auxiliaries per 500 words, as explained in Chapter 4.8.2.2. In year 2 the unit was represented by a literature essay of approximately 1000 words in length; in year 4, by a dissertation of between 5000 and 7000 words. Both exercises were written in the students' own time. The items which were *counted* at this preliminary stage were the 3pers sing form of *werden*

and of *sein, wird* and *ist* respectively.[2] The results of the raw *count* are shown in Table 5.1.

	wird		Total *wird*	*ist*		Total *ist*
	Year 2	Year 4		Year 2	Year 4	
A8	9	7	16	0	1	1
A7	8	15	23	3	10	13

Table 5.1: *wird* and *ist* passives, Years 2 and 4; 2 respondents

5.2.2   Quantitative Concordance Part 2

Since the initial raw score calculations were based on units of production of an *approximate* length (between 5000 and 7000 words) and therefore lacked precision, it was decided to make a more specific calculation of instances of the passive-per-500-words, a measurement which was introduced in Chapter 4.8.2.2.   Therefore, an electronic *count* was carried out to establish the precise number of words in each of the written texts. The total number of passives-per-500-words calculated by this means is indicated in Table 5.2.

	A8			**A7**		
	Word count	*wird*/500	*ist*/500	**Word count**	*wird*/500	*ist*/500
Year 2	797	5.6	0	1236	3.2	1.2
Year 4	5192	0.6	0.1	6825	1.1	0.7

Table 5.2: passive types/tokens per-500-words, Years 2 and 4; 2 respondents

A comparison of Tables 5.1 and 5.2 indicates the refinements which are provided by a more precise calculation. As the text length increases, so an exponentially similar increase in the number of passive tokens might be expected.[3] However, the figures in Table 5.2 show that A7's production of both *wird* and *ist* passives has fallen between years 2 and 4, and that only in the case of A8's *ist* passives has there been an increase in production. The results also indicate that in both years the L2 G learner (A8) produced

---

[2] See Table 5.5 for a complete list of the item types which were the subject of the second phase investigation.
[3] But note Richards' comment reported in Chapter 4.8.2, suggesting that as learners become more fluent and are able to produce an increased number of structures, token counts may decrease.

fewer *sein* passives than the L1 G user (A7). As for *werden* passives, both respondents produced fewer of them in year 4 than in year 2, with the drop by A8 being greater than that of A7.

5.2.3 Extension of the Preliminary Analysis

The results of the preliminary analysis were deemed of sufficient interest to extend the study, since the expectation of an overall increase in the number of passive tokens over time had not been realised. Concordances of the same items from units of production from the Year 1 writing of the same two respondents were undertaken. In year 1, the unit is represented by a short essay of approximately 300 words describing a proposal by a tourist bureau for a weekend trip. While the length of this text is short, it is indicative of the type of exercise undertaken by learners at first year level, and since, according to the results shown in Table 5.2, a NS produces both *werden* and *sein* passives at this stage in a pattern which differs from a NNS both numerically and typologically, it is useful to include this exercise in order to be able to carry out both a diachronic and a synchronic analysis of two language types.

	wird			Total *wird*	*ist*			Total *ist*
	Year 1	Year 2	Year 4		Year 1	Year 2	Year 4	
A8	8	9	7	24	0	0	1	1
A7	0	8	15	23	0	3	10	13

Table 5.3: *wird* and *ist* passive tokens, Years 1, 2 and 4; 2 respondents

The conversion to passives-per-500-words was then applied to the figures in Table 5.3 resulting in the figures indicated in Table 5.4.

	A8			A7		
	Word count	*wird*/500	*ist*/500	Word count	*wird*/500	*ist*/500
Year 1	280	14.2	0	352	0	0
Year 2	797	5.6	0	1236	3.2	1.2
Year 4	5192	0.6	0.1	6825	1.1	0.7

Table 5.4: *wird* and *ist* passive tokens per-500-words, Years 1, 2 and 4; 2 respondents

An examination of the figures from all six assignments lead us to make the following observations: firstly, A8 produced more instances of the *wird* passive-per-500-words in her shortest (Year 1) text than in either of the others and the number of instances per-500-words decreased over three years. This could be the effect of the learner's increasing skill in use of the passive, with less reliance on the basic periphrastic forms. On the other hand, it might be an indicator of passive avoidance. Secondly, in stark contrast to A8, A7 produced no *wird* passives in year 1. Thirdly, as we might expect given the nil return for *ist* passives in year 2, A8 produced none in year 1 either, and even in year 4 the level of production is very low. Fourthly, A7's production exhibits an increase for both *wird* and *ist* passives in the second year exercise but a decrease by year 4. We can conclude that no pattern emerges which is common to both respondents, except in so far as neither respondent produced an *ist* passive in year 1. The issue of changes in the use of the passive being dependent on text style is considered in Section 5.3.1. It should also be borne in mind that it is not possible to deduce directly from the student's written production whether the non-appearance of an item, either at this early stage of the analysis or in subsequent stages, is due to a conscious decision on the part of any of the learners to avoid the passive. The limitations in the scope of a learner corpus in this respect will be discussed in Chapter 7.

5.3   Quantitative Analysis: A corpus

Following the preliminary stage described in section 5.2, the analysis moved to a quantitative cross-sectional study of passive typology produced in the writing of the entire A cohort, from year 1 to year 4, the results of which are reported in section 5.3.1. The process of converting the raw scores to passives-per-500-words will be used again in the assessment of longitudinal change in the students' production in section 5.3.2. The raw token scores are presented first because we are focusing on the frequency *count* of specific tokens and not on the broader question of typological frequency (i.e. the frequency of *werden* or *sein*). As was pointed out in the introduction to this chapter, the results of the *count* and *sort* processes will be discussed in Chapters 5, 6 and 7; in this Chapter 5, each illustrative table is followed by a number of comments which indicate noteworthy variation across the cohort, in keeping with the overarching research question. A *comparison* of the students' correct matching of form and function of the *werden* and *sein* passives will provide the basis of the qualitative evaluation which

follows in Chapter 6 and subsequently the three *count*, *sort* and *compare* processes will provide evidence for the discussion of passive use by NNSs and NSs in Chapter 7.

5.3.1 Cross-sectional Analysis

The lack of evidence of any commonality in the development of production of two passive forms by two respondents encouraged us to expand the corpus size and re-focus the project. Research questions 1 and 2 provided the focus for the examination of the written output of all the A cohort over three years. The data sample from year 2 was changed from one literature essay written without time constraints to two timed examination answers in the form of a response to a question (as opposed to either guided composition or formal grammar exercises), therefore reflecting more closely Granger's definition of a learner corpus (Granger 2002: 124) which was identified in Chapter 2.6.3.1. Including these data enabled the examination of written samples produced in more than one language setting, both under time constraints and as free writing exercises. Furthermore, the additional, alternative data were readily accessible (after the production of electronic files) for an additional respondent, resulting in a larger corpus with a total of four exercises per each of eight students plus three exercises for a further student. It was expected that there would be some notable differences between the timed and free-writing exercises, if, for example, learners were unable to employ strategies to monitor their grammatical accuracy while constrained by the time limits imposed in an examination.[4]

In this stage of the analysis, the number of separate items to be *counted* was increased from two to nine, in order to extract examples of all the morphological third person singular and plural (3pers sing and 3pers pl) forms of the auxiliaries *werden* and *sein*.[5] The items are listed in Table 5.5. These were entered into the concordancer for each respondent, in order in the first instance to *count* and then *sort* the various forms of the auxiliaries before linking them to the past participles with which they combine to represent syntactic passives. The grammatical forms and use of the tenses in the construction of the passive in German have been explained in Chapter 3.

---

[4] See Chapter 2, footnote 21.
[5] Only 3pers forms were extracted; other forms were not considered in this investigation because of an anticipated nil occurrence of personal references, i.e. first and second person forms, in the exercises in the corpus. Students are advised to avoid these in academic writing unless specifically requested to include them.

It should be noted that there was no particular requirement to produce passives in any of the exercises in the corpus,[6] although since the second examination question is formulated using a passive, that construction might be expected to occur more frequently if the question serves as a prime for the respondents to produce a text in the same genre. In all of the exercises, the learners themselves made the decision regarding the structures to use.

Item	Type	Grammatical form
1	*wird*	3pers sing pres *werden*
2	*ist*	3pers sing pres *sein*
3	*werden*	3pers pl pres *werden*
4	*sind*	3pers pl pres *sein*
5	*wurde*	3pers sing imp *werden*
6	*war*	3pers sing imp *sein*
7	*wurden*	3pers pl imp *werden*
8	*waren*	3pers pl imp *sein*
9	*worden*	PP *werden* contracted form of *geworden*

Table 5.5: Search items entered in the concordancer

5.3.1.1 Year 1

	A1	A2	A3	A4	A5	A6	A7	A8	A9	Total
Total words	311	305	300	N/A	310	278	352	280	278	**2,414**

Table 5.6: Word counts, Year 1 exercise
N/A = no data available

Tables 5.7 (*werden* passives) and 5.8 (*sein* passives) indicate the number of tokens of the selected passive types (Table 5.5) retrieved from the year 1 written exercise. Rather than showing totals only, the *count* of individual items is retained in order to demonstrate the variety or otherwise of types used by the learners. Although the word counts are low (Table 5.6), the variation for individual production of *werden* passives between zero and 13 is worthy of note, as is the difference between the use of *werden* and *sein* passives in almost every case, and the fact that only A3 produced no passives at all.

---

[6] A grammar cloze test where a learner was required to fill gaps with a passive construction is an example of the type of exercise where only a passive would be acceptable.

Two comments should be noted regarding the composition of Tables 5.7 and 5.8: firstly, no data are available for A4 for the year 1 exercise (indicated by N/A) and secondly, the production of A7 is included both at this point in the process in order to compare her production with that of her contemporaries with L1 E in the A cohort and also with the sample of L1 Gs in Section 5.4.

Item	Type	A1	A2	A3	A4	A5	A6	A7	A8	A9	Total
1	*wird*	1	4	0	N/A	0	0	0	8	0	**13**
3	*werden*	0	2	0	N/A	1	3	1	4	3	**14**
5	*wurde*	0	0	0	N/A	6	1	2	0	0	**9**
7	*wurden*	0	0	0	N/A	1	0	0	0	0	**1**
9	*worden*	0	0	0	N/A	0	0	1	1	0	**2**
	TOTAL	**1**	**6**	**0**	**N/A**	**8**	**4**	**4**	**13**	**3**	**39**

Table 5.7: *werden* passive tokens, Year 1; 8 respondents

Item	Type	A1	A2	A3	A4	A5	A6	A7	A8	A9	Total
2	*ist*	1	0	0	N/A	0	0	0	1	0	**2**
4	*sind*	0	0	0	N/A	0	0	0	0	1	**1**
6	*war*	0	0	0	N/A	2	0	0	0	0	**2**
8	*waren*	0	0	0	N/A	0	0	0	0	0	**0**
	TOTAL	**1**	**0**	**0**	**N/A**	**2**	**0**	**0**	**1**	**1**	**5**

Table 5.8: *sein* passive tokens, Year 1; 8 respondents

Remembering that the production of a larger number of *werden* passives than *sein* passives is to be expected given the order of input in the learning environment and the fact that Duden (1998) attests to a lower number of *sein* passives than *werden* passives in its corpus (see Chapter 3.5), across the cohort we *count* 39 passive tokens with auxiliary *werden* (all five types) and five with auxiliary *sein* (two types). Whatever the final state of the production of the passive by the A cohort might be, this sample is certainly a reflection of the current state of the construction in their interlanguage. It is possible that this current state will remain unchanged; we can take note of changes in this respect during the longitudinal analysis of the students' production over three years.

The present tense would be introduced first in an instructional setting (e.g. Buscha & Szita 2007: 221). This order of instruction perhaps accounts for the fact that of the two present tense types, types 1 (*wird*) and 3 (*werden*) are produced equally across the sample, the latter by more respondents than the former. The learners may be producing these forms as formulaic sequences as described in Chapter 2.5.3 6. Present tense items 2 and 4 for *sein* are produced in a limited way by four of the eight participants. It is, however, the case that the imperfect of *werden* appears in the greatest numbers with just one individual, A5.
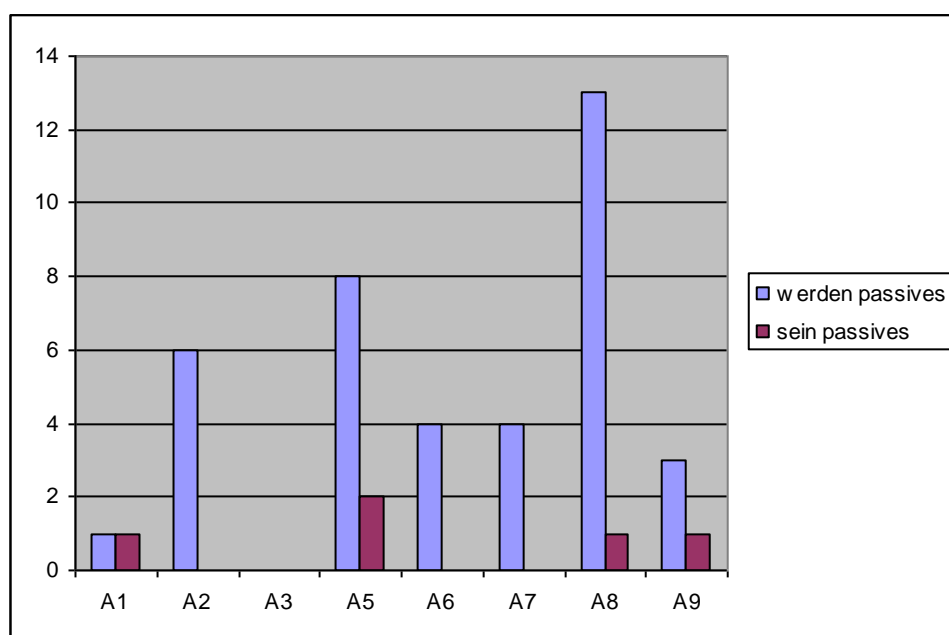
Figure 5.1: Total passive types, Year 1; 8 respondents

Figure 5.1 shows the production of the passive in year 1 by the A cohort in diagrammatic form.

Regarding individual respondents, while A7 produces no *sein* passives in year 1, her production of *werden* passives is not limited to present tenses, in contrast to the examples from the other members of the A cohort. It would appear from Figure 5.1 that A7's production of the passive equates exactly to that of A6 in year 1 but it is noted that this figure displays only a numerical frequency *count*. A more accurate comparison will be obtained in Section 5.3.2, the longitudinal analysis, based on the passives-per-500-words calculation. The only other similarity between A7 and the rest of the A cohort is

the lack of *sein* passives in the production of A3, who produces no passives at all at this stage.

A8 is the most prolific user of the passive with the auxiliary *werden.* A3 produces no tokens for *werden* and none for *sein* either. In the case of the auxiliary *sein*, there are only five examples across the entire dataset, two of which are produced by student A5 who is to be regarded as the respondent displaying the most frequent use of the passive, albeit with such a limited number of tokens. Will these patterns be repeated in the written production of these students over the following years' exercises? If we categorise A5 and A8 as the respondents with the most wide-ranging use of the passive at four types each in year 1, may other respondents' production in other exercises demonstrate similar use? These questions among others will be addressed in Chapter 6. At this stage of the analysis, establishing the frequency of passive use by these learners by means of a token *count* is important in order to be able to calculate the learners' accuracy of choice between the *werden* and *sein* passives, since it is that accuracy which is of particular interest in this study.

5.3.1.2 Year 2

Exam question 1

The precise number of words in each corpus text is shown in Table 5.9. Tables 5.10 and 5.11 indicate the number of passive tokens and types occurring in each text. These two tables show only the item types which were produced. Zero production of a type is not recorded for the examination answers because of the notable absence of passives, particularly in question 1 (a total of 11 overall), but a *comparison* between all the learners can be drawn from Figure 5.2.

	A1	A2	A3	A4	A5	A6	A7	A8	A9	Total
Total words	251	185	246	282	241	261	261	241	239	**2207**

Table 5.9: Word counts, Year 2, exam question 1

Item	Type	A1	A2	A3	A4	A5	A6	A7	A8	A9	Total
1	*wird*	0	0	0	1	0	0	2	0	0	3
3	*werden*	0	0	0	0	1	0	2	2	1	6
5	*wurde*	0	0	0	0	1	0	1	0	0	2
	TOTAL	0	0	0	1	2	0	5	2	1	11

Table 5.10: *werden* passive tokens, Year 2, exam question 1; 9 respondents

Item	Type	A1	A2	A3	A4	A5	A6	A7	A8	A9	Total
4	*sind*	0	0	1	1	0	1	0	0	1	4
	TOTAL	0	0	1	1	0	1	0	0	1	4

Table 5.11: *sein* passive tokens, Year 2, exam question 1; 9 respondents

It can be seen that the raw frequency *count* of passive tokens in this examination exercise is much lower than in the Year 1 exercise which was produced in the students' own time, which is to be expected since the Year 2 texts are shorter. Across the cohort, we *count* only 11 passive tokens with auxiliary *werden* (three different types) and four with auxiliary *sein*. Production of *sein* passives is limited to a single form, 3pers pl pres (*sind*). The present tense of the auxiliaries is used almost exclusively apart from two instances of item 5 (*wurde*).

In this exercise, A7's production is limited to *werden* passives, but she produces the most passive tokens. Her production of *sein* passives remains at zero, as it was in year 1. The two most prolific users of the passive in the year 1 exercise, A5 and A8, are joined by A9 in producing two tokens each of the nine searched items. Two respondents (A1 and A2) exhibit no instances of passive in this exercise.
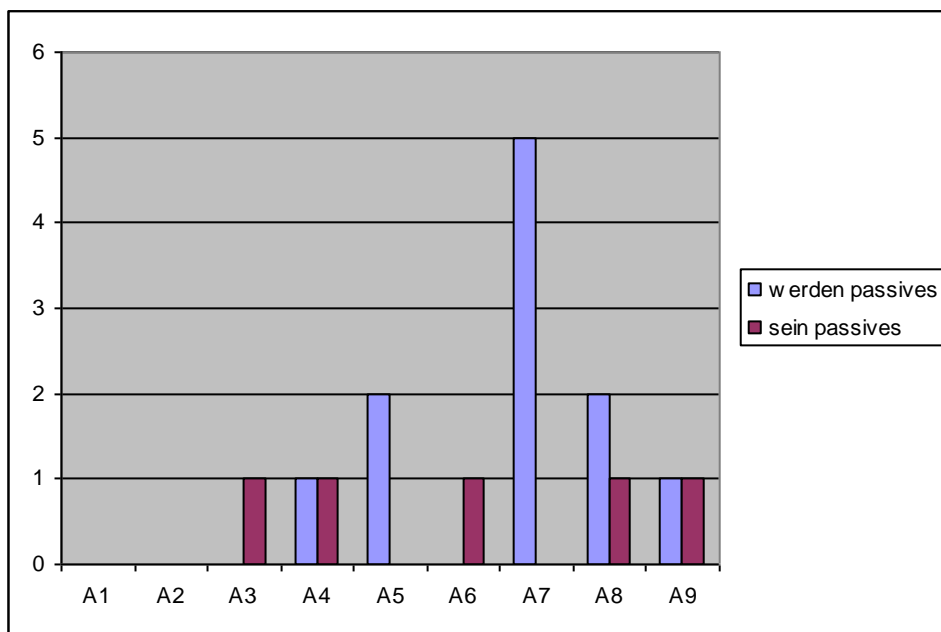
Figure 5.2: Total passive types, Year 2, exam question 1; 9 respondents

Exam question 2

The precise number of words in each corpus text is shown in Table 5.12. Tables 5.13 and 5.14 indicate the number of passive tokens and types occurring in each text. Once again, as with examination question 1, zero results for types are not shown in the tables and a comparison between all the respondents may be observed in Figure 5.3.

	A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**Total**
Total words	178	183	167	151	214	182	143	176	170	**1564**

Table 5.12: Word counts, Year 2, exam question 2

Item	**Type**	**A1**	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**Total**
1	*wird*	1	5	0	0	4	0	6	0	0	**16**
3	*werden*	0	3	0	1	2	0	2	3	0	**11**
7	*wurden*	0	0	0	1	0	0	1	0	0	**2**
9	*worden*	0	1	0	0	0	0	0	0	0	**1**
	TOTAL	**1**	**9**	**0**	**2**	**6**	**0**	**9**	**3**	**0**	**30**

Table 5.13: *werden* passive tokens, Year 2, exam question 2; 9 respondents

133

Item	Type	A1	A2	A3	A4	A5	A6	A7	A8	A9	Total
2	*ist*	1	1	0	0	0	3	1	0	0	**6**
4	*sind*	0	0	4	1	2	2	2	0	4	**15**
	TOTAL	**1**	**1**	**4**	**1**	**2**	**5**	**3**	**0**	**4**	**21**

Table 5.14: *sein* passive tokens, Year 2, exam question 2; 9 respondents

This examination exercise was expressed in the form of a question: '*Wie wird das Europäische Parlament gewählt*?' (= How is the European Parliament elected?). The two examination answers were included in the corpus in order to extend the exercise type to writing produced under time constraints as well as without, in line with the definition of a learner corpus identified by Granger (2002: 124). Since the passive is used in the formulation of the second examination question, and very possibly acts as a prime, it would be expected that answers would also exhibit use of the construction, resulting in a text with no attribution of agent. This expectation is clearly justified by the notable increase in the number of passive tokens in this exercise compared to exam question 1, particularly in view of the fact that the overall word count for the second question was in most cases lower than the first. However, it is noted that A3, A6 and A9 produce no *werden* passives at all and A1, only one.

Across the cohort, 30 passives are *counted* with auxiliary *werden* and 21 with auxiliary *sein*. All the respondents attempt the passive to some extent in this exercise, with A1 producing the fewest tokens overall (2 in total, one of each 3pers sing type). A7 is the most prolific user of the passive, followed by A2 with 9 *werden* and 1 *sein* and A5 with 6 *werden* and 2 *sein*. For the first time, A7 produces *sein* passives, in the exercise where an increase in the production of passives is expected.

It should be noted that A5 was identified as one of the respondents with the most wide-ranging use of the passive following the analysis of the Year 1 exercise, and also figured in the Year 2 first examination question as one of the most prolific producers of the passive, even though on a very limited scale. In this exercise it is A2 who, with 10 tokens across four types, demonstrates the widest variation in production. Contrary to predictions, A8's production of the passive in this exercise barely rises above that of the first examination question and drops in comparison to the Year 1 exercise.

Again, present tenses of the auxiliaries predominate, with the most commonly used form being the 3pers sing pres of *werden*, with a total of 16 tokens, followed by the 3pers pl pres of *sein* (15 tokens) and the 3pers pl pres of *werden* (11 tokens). The single example of a past tense is no exception since it is formed with PP combined with present tense *sein*.

Figures 5.2 and 5.3 show the production of the passive by the A cohort in the Year 2 examination questions in diagrammatic form.
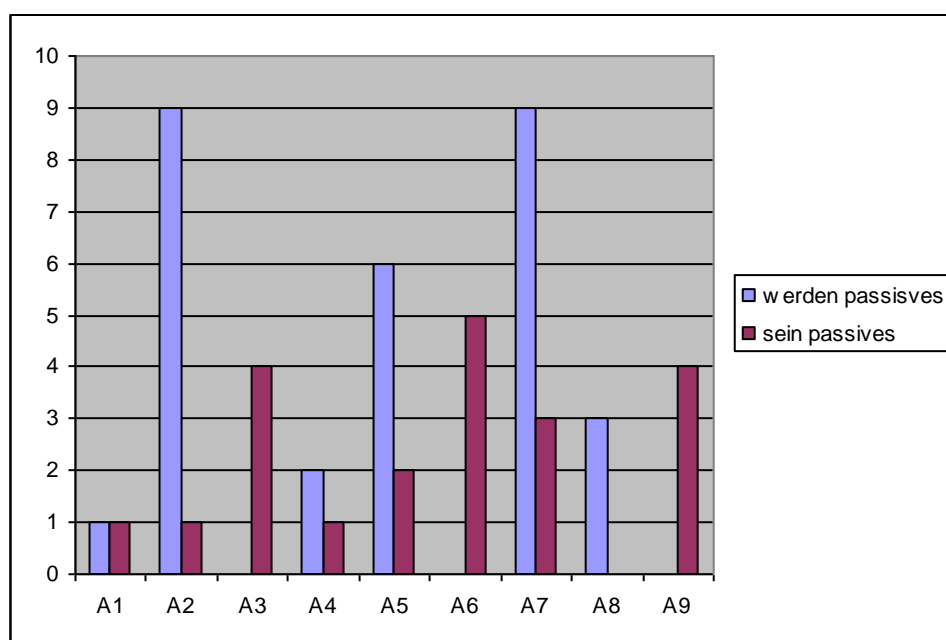
Figure 5.3: Total passive types, Year 2, exam question 2; 9 respondents

5.3.1.3 Year 4

The precise number of words in each corpus text is shown in Table 5.15. Tables 5.16 and 5.17 indicate the number of passive tokens and types occurring in each text.

	A1	A2	A3	A4	A5	A6	A7	A8	A9	Total
Total words	4319	5592	5050	4440	6121	4683	6825	5192	6037	**47,719**

Table 5.15: Word counts, Year 4

Item	types	A1	A2	A3	A4	A5	A6	A7	A8	A9	Total
1	*wird*	20	15	1	6	11	5	21	7	35	**121**
3	*werden*	22	23	1	9	27	9	48	17	15	**171**
5	*wurde*	4	3	1	0	7	0	26	20	0	**61**
7	*wurden*	0	2	0	1	7	0	7	8	0	**25**
9	*worden*	0	4	1	1	3	1	4	4	0	**18**
	TOTAL	46	47	4	17	55	15	106	56	50	396

Table 5.16: *werden* **passive tokens, Year 4; 9 respondents**

Item	types	A1	A2	A3	A4	A5	A6	A7	A8	A9	Total
2	*ist*	2	0	4	7	1	1	13	1	6	**35**
4	*sind*	7	5	5	12	4	2	9	1	6	**51**
6	*war*	8	0	4	1	0	1	4	0	1	**19**
8	*waren*	5	0	4	6	1	1	2	1	1	**21**
	TOTAL	22	5	17	26	6	5	28	3	14	126

Table 5.17: *sein* **passive tokens, Year 4; 9 respondents**

All respondents except A3 and A4 produce more *werden* passives than *sein* passives in the Year 4 exercise. It is interesting that these latter two cases represent a reversal of Duden's statement that *werden* passives predominate over *sein* passives in contemporary German texts. Even though the learners' texts cannot be directly compared with Duden's corpus in terms of size and genre, the production of A3 and A4 is clearly different from the other learners in this respect. This point will be considered further in Chapter 7. Present tenses are predominantly used across the cohort, with A7 and A8 being the exceptions in additionally producing a high number of item 5 (*wurde*). A7 produces the highest number of *werden* passives overall with her tally of 106 tokens, and it is worth noting again here that this is the student in the A cohort with L1 G. In the remainder of the A cohort, A8 and A5 produce only half as many *werden* passives, 56 and 55 respectively. A5 continues to demonstrate the wide-ranging production she has shown in all of the exercises analysed.
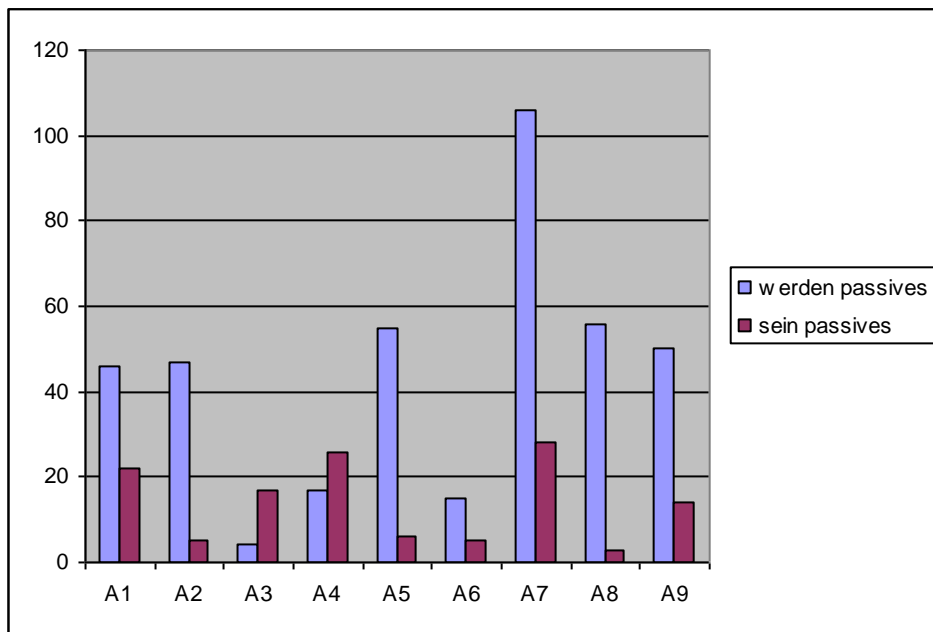
Figure 5.4: Total passive types, Year 4; 9 respondents

Figure 5.4 shows the production of the passive by the A cohort in the Year 4 exercise in diagrammatic form.

In year 4, although present tenses are used frequently by A7, she produces all the search items to some extent, which is not the case for the other members of the A cohort where at least one zero is recorded for each individual (Tables 5.16 and 5.17).

By this stage, all the A respondents are producing both types of passive. A discussion of the variation in usage of the passive by NSs and NNSs follows in Chapter 7.

Figures 5.1-5.4 show the results of the CIA which has been carried out *comparing* the IL of a number of respondents, eight of whom have the same mother tongue. The contributions of the ninth member of the A cohort, A7, indicate that her position as a NS of German sets her apart from her contemporaries in the A cohort in terms of the production of passives. Her written production in the year 4 exercise will be considered again with reference to four members of the B cohort in Section 5.4.

5.3.2   Longitudinal Analysis

Returning to research question 2, the data were then subjected to a longitudinal analysis. Research question 2 is expressed in the following way:

Do these forms (of the passive) vary over time?

Table 5.18 offers a numerical, quantitative comparison between the numbers of tokens of the types of passive in this survey which the L1 E learners of German produced over 4 years of study. It also includes the data sample of A7, the NS of German, in order to compare the longitudinal variation in her use of the passive with that of her contemporaries in the A cohort. The ratio conversion of passives-per-500-words is reintroduced at this juncture in order to enable a more precise evaluation of similarities and differences in the production of the individuals as an extension of the type/token *count* which was provided by the cross-sectional analysis.

It was predicted that there would be a greater number of passive tokens in the answer to examination question 2 than in the other exercises, with the passive in the question being used as a model by the respondents. This prediction is seen to be justified in the case of approximately half of the cohort except A3 (*werden*), A6 (*werden*), A8 (*werden* and *sein*) and A9 (*werden*). The greatest number of *werden* passives-per-500-words across the whole dataset appears in the examination question 2 exercise of A7 (31.5), setting her apart once again from her contemporaries with L1 E. It is interesting to note that the incidence of *werden* passives decreases significantly in most cases under time constraints in examination question 1 perhaps because under the pressure of time there is a reluctance on the part of the learner to use more complex constructions. No such conclusion can be drawn from the analysis of *sein* passives in this exercise.

If the results of the year 2 exercises are disregarded because they are produced under time-constrained conditions unlike the year 1 and year 4 exercises, and respondent A4 for whom no data are available for year 1 is excluded, a linear increase from year 1 to year 4 in *werden* passives by A1 and A3 is noted, and in *sein* passives by A1, A2, A3, A6 and A7. A3 produced no *werden* passives at all until year 4. Her production of *sein* passives remains relatively stable over three exercises but reaches the highest total in the entire dataset in the Year 2 examination question 2.

Respondents A5 and A8 have merited comments in the preceding cross-sectional analysis. In this longitudinal phase the production of the former's *werden* passives remains relatively stable over Year 1 and Year 2 exam question 2 and between Year 2 question 2 and Year 4. In comparison, A8's elevated total in year 1 is reduced to

exactly the same frequency as A5 in examination question 1; even in examination question 2 where a greater number of passives were predicted A8's total does not reach the same level as in year 1. Only A5, A8 and A9 produce fewer passive tokens of both types from year 1 to year 4.

The production of student A7 merits further comment in this longitudinal analysis because it is assessed at this juncture against the production of the other eight members of the same student cohort. It will also be considered with relation to that of the members of the NS German cohort, as stated in Section 5.3.1.1.

From year 1 to year 2, an increase in A7's production of *werden* passives can be seen, followed by a slight drop in year 4, if we discount examination answer 2 which may represent the result of priming in the question in a similar way to the rest of the A cohort, as suggested in Section 5.3.1. Similarly, from a nil response in year 1 and the first examination question, the high level of production of *sein* passives by A7 in examination question 2 can also be explained by the exercise type.

Compared to the rest of the A cohort, only A1 and A3 similarly produce more *werden* passives by year 4 than they did in the earliest stage. There appears to be no common pattern; each of these two displays particular variables. In the case of A1, production of both types of passive is exactly the same until year 4. A3 relies only on *sein* passives until year 4 as noted earlier. She and A4 are the only two respondents in the A cohort to produce more *sein* passives than *werden* passives in year 4. In addition, it is impossible to draw convincing conclusions in respect of a comparison with A7 from A4's incomplete data sample. No conclusion can be drawn from a comparison of A7 with the other members of the A cohort in terms of NS versus NNS usage of the passive.

	A1		A2		A3		A4		A5		A6		A7		A8		A9	
	w	s	w	s	w	s	w	s	w	s	w	s	w	s	w	s	w	s
Year 1	1.6	1.6	9.8	0	0	1.6	N/A	N/A	13	3.2	7.1	0	5.7	0	23.2	1.7	5.4	1.8
Year 2 exam question 1	0	0	0	0	0	2	1.7	1.7	4.1	0	0	1.9	9.6	0	4.1	0	2.1	2.1
Year 2 exam question 2	2.8	2.8	24.5	2.7	0	12	6.6	3.3	14	4.7	0	13.7	31.5	10.5	8.5	0	0	11.8
Year 4 dissertation	5.3	2.5	4.2	0.4	0.4	1.7	1.9	2.9	4.5	0.5	1.6	0.6	7.8	2.1	5.4	0.3	4.1	1.2

Table 5.18: Passives-per-500-words; A corpus, 9 respondents

w = *werden*

s = *sein*

N/A = No data available

There is no robust evidence in these results for fossilization in the production of the particular learners in this data sample, since there is significant fluctuation in the production of every individual.

In Chapter 6, the information in Table 5.18 is scrutinised from the point of view of correctness of choice between *werden* and *sein* passives, with a view to evaluating this important element in the way the German passive is used by advanced learners with L1 E. In its approach to research question 4, the aim in that chapter will be to suggest possible reasons for learner choice.

5.4   Quantitative Analysis: B corpus

The written production of this small group of four L1 G respondents is used in this study as examples of contemporary German texts because they are NS users of German who will presumably have been exposed to more academic writing and speech in German than the learners in the A cohort and therefore should be more proficient in the usage of complex constructions. While this may not necessarily be the case, their writing skills as L1 G users will undoubtedly be more advanced than those of the L1 E learners of German, and therefore the B corpus texts can serve as a *comparison* to those in the A corpus.

Student A7's production is examined in this section in her special case position as a member of the A cohort but with L1 G in order to remark on similarities and differences between NS and NNS usage of the passive. Having regarded her Year 4 production as an example of NS writing *compared* to that of her A cohort contemporaries in Section 5.3.2, it is now *compared* with that of L1 G users in the B cohort who have been studying in an entirely NS environment. The production of the B cohort is regarded as a suitable dataset for *comparative* purposes since the texts were written by undergraduate students who produced the dissertations or long essays for assessment at a similar stage in their academic careers to that of the L1 E students in the A cohort, as explained in Chapter 4.6.1.2.

At this stage of the process, considerations returned to research question 3, namely:

Which forms of the passive are evident in the written production of adult NSs of German?

Since one single exercise was collected from the four respondents in the B cohort, only a cross-sectional analysis could be carried out. The results of the numerical *count* and *sort* process are shown in Tables 5.19 and 5.20. These two tables include the production in year 4 of the lone L1 G user in the A cohort, which is similar in text type to, and can therefore be *compared* with, the B respondents' production, since they all comprise a longer length dissertation-type exercise which demonstrates a student's research skills and contributes to the student's final assessment.

Item		B1	B2	B3	B4	A7	Total
1	*wird*	24	48	27	23	21	**143**
3	*werden*	65	57	91	32	48	**293**
5	*wurde*	14	5	18	2	26	**65**
7	*wurden*	14	3	17	0	7	**41**
9	*worden*	2	2	11	0	4	**19**
	TOTAL	119	115	164	57	106	561

Table 5.19: *werden* passive tokens, B corpus; 5 respondents

Item		B1	B2	B3	B4	A7	Total
2	*ist*	14	17	14[7]	1	13	**59**
4	*sind*	12	15	7	3	9	**46**
6	*war*	1	1	3	0	4	**9**
8	*waren*	0	0	0	0	2	**2**
	TOTAL	27	33	24	4	28	116

Table 5.20: *sein* passive tokens, B corpus; 5 respondents

---

[7] One instance was found in the B corpus of a compound past tense *sein* passive i.e. ist+PP+gewesen. It is the only example of its kind and can be seen as an indicator of the more complex language structures of which L1 G users are capable.

As demonstrated in Tables 5.19 and 5.20, in all cases the incidence of *werden* passives is greater than that of *sein* passives, in line with Duden's (1998) findings. Present tenses predominate, although all of the search items are present to some extent.

	Word count	w	Per 500	s	Per 500
B1	9183	119	6.5	27	1.5
B2	11561	115	5	33	1.4
B3	11414	164	7.2	24	1.1
B4	1963	57	14.5	4	1.0
A7 year 4	6825	106	7.8	28	2.1

Table 5.21: Passives-per-500-words, B corpus

w = *werden*
s = *sein*

Comparing the number of passives-per-500-words produced by A7 with the B corpus, we note some variation in production of *werden* passives between all the respondents in Table 5.21. There is a broad similarity between all the B respondents in terms of their production of the *sein* passive, with A7's production of that type being slightly higher.
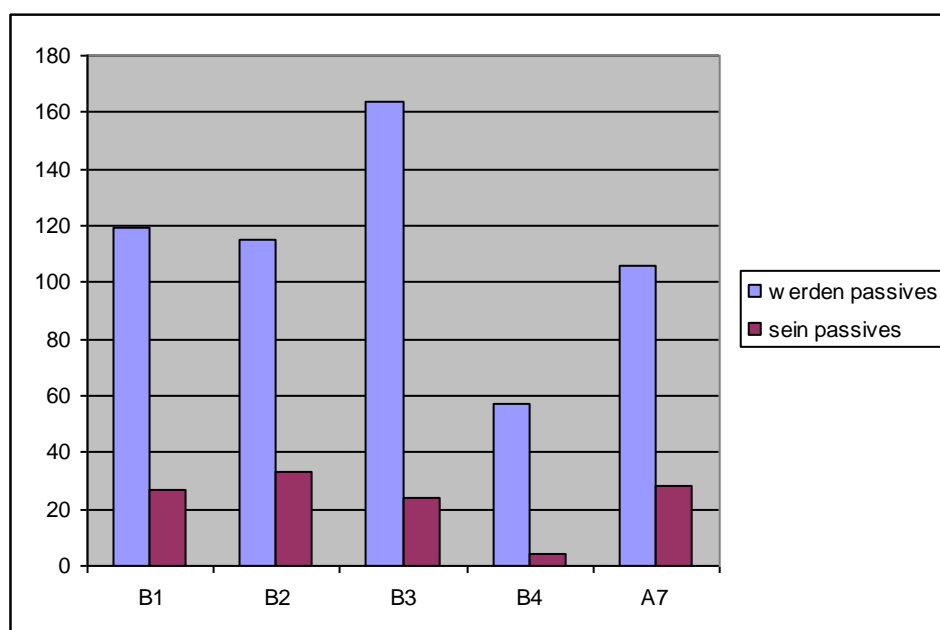
Figure 5.5: Total passive types, B corpus; 5 respondents

While it was not possible to draw any conclusions in the longitudinal analysis regarding similarities between A7's production and that of L1 E learners of German, a comparison of Figure 5.5 with Figure 5.4, which shows production across the A cohort in an exercise similar to the one contributed by the B cohort, indicates by including A7 in

both that A7's production profile is closer to the production of others with L1 G than it is to those in the L1 E cohort. This is true both overall and in respect of the predominance of *werden* passives over *sein*. A discussion of A7's position in relation to both the A and the B corpora follows in Chapter 6.

5.5   Conclusions Drawn from Chapter 5

The results of the analyses of the L1 E corpus exercises from Years 1 - 4 indicate that the passive with *werden* is more common in the production of most of this cohort than the passive with *sein* and that the present tenses of the auxiliaries are favoured. There are noticeable differences in the cross-sectional production of individual users and although an increase in use of a construction might be expected over time, the results show that this is not always the case. As we shall see in Chapter 6, longitudinal variation as well as cross-sectional variation is observable. Such variation constitutes one of the points for discussion, in order to account, for example, for a large number of *werden* passives (A8, year 1) or *sein* passives (A6, examination question 2). When compared with the passive constructions produced by L1 users of German, it may be that variation between and within the L1 E learners' production simply reflects a stage of interlanguage development.

The qualitative analysis in Chapter 6 will reveal whether or not the L1 E learners make the correct choice of auxiliary when it comes to opting for one or other of the periphrastic forms of the German passive. This analysis will add a further dimension to the exploration of frequency *counts* which has been described in Chapter 5. A *comparison* will be made with the choices of the L1 G users. In Chapter 7, a closer examination of the NSs' and NNSs' use of the passive is undertaken with the aim of reaching some conclusions relating to nativelike or non-nativelike use of the passive construction by L1 E learners.

CHAPTER 6 - QUANTITATIVE AND QUALITATIVE ANALYSIS: THE FORM/FUNCTION RELATIONSHIP IN THE GERMAN PASSIVE

6.1.   Introduction

In Chapter 3, the multiplicity of German passive forms and the complex nature of its semantics were described. In Chapter 6, we turn to an assessment of the usage of this construction exhibited in the written production of the learners and users in the LINCS corpus. While Chapter 5 provided a quantitative assessment of passive typology in the LINCS corpus by *counting* and *sorting* data, Chapter 6 begins with a different type of quantitative *count* which will enable a *comparative*, qualitative analysis to be carried out on the three subsections as described in Chapter 4.6, the A corpus, the B corpus and the C corpus. A response will be given to research question 4 with reference to the answers derived from questions 1, 2 and 3 during the frequency *count* which was described in Chapter 5. The four research questions are repeated here for the sake of clarity.

1. Which forms of the German passive are produced in written interlanguage by adult learners with L1 E?
2. Do these forms vary over time?
3. Which forms of the passive are evident in the written production of adult NSs of German?
4. What similarities or differences can be observed between the use of the passive by NSs and NNSs in the respective subcorpora?

Having provided a response to research questions $1 - 4$, it will then be possible to return to the general aim of the study, that is, to investigate patterns in the usage of the German passive by a cohort of learners with L1 English at different stages in their learning process in the formal learning environment that was described in Chapter 4.6 and to consider what explanations could be advanced to account for these patterns.

Ellis & Barkhuizen suggest that 'there is no one way of doing qualitative research' (Ellis & Barkhuizen 2005: 254) and bearing this point in mind, Chapter 6 contains a

description of a combination of methods: a quantitative, formula-based calculation of a qualitative evaluation of the learners' use of the correct syntactic passive for the semantic context they have chosen. The rationale for using a formula at this stage of the analysis is based on Ellis & Barkhuizen's remark that '[t]o tell the full story [of learners' acquisition of language] we need to consider what learners get right as well as what they get wrong' (Ellis & Barkhuizen 2005: 73). The formula adopted, as described in Section 6.2, provides a means of measuring the accuracy of syntactic passive use in contexts chosen by the learners themselves. It does not take account of learner behaviour, for example avoidance strategies. This point will be addressed again in Chapter 7. Following the introduction of the formula, an overview is given in diagrammatic form of the quality of production of the relationship between form and function as revealed by evidence from the corpus. 'Quality' in this instance refers to the correctness of choice made by the learners in determining whether to produce an actional or a statal passive.

The data for the A respondents are firstly presented globally in the form of figures indicating the passive typology and correctness of the choices made by the learners. This overview is followed by a more detailed examination of the data for each respondent in turn in a series of mini case study reports. These are not full case studies as the term is described in Chapter 4.8.1, but the goal here is to offer an insight into interesting aspects of the learners' choices. In this format, the respondents' production of the passive is scrutinised longitudinally, comparing passive intent over four written exercises. By 'passive intent' we mean that the learner was assumed by the analyst to be attempting to produce a passive. This was explained in Chapter 4.4.4 and further elaborated in Chapter 4.8.2.1 and Chapter 5.1.

The chapter continues with a section in which evidence of the use of alternatives to the *werden* and *sein* passives is discussed and reference is then made to the sample of input data in the C corpus. The final section draws conclusions from the findings of the assessment of the form/function relationship of the German passive as it is produced by the subgroups of the LINCS corpus.

6.2   Target-Like-Choice Analysis

Pica (1983:70) coined the term 'Target-Like-Use Analysis' as an alternative way of assessing the accuracy of morpheme acquisition in English as a Second Language (ESL). Her formula requires the calculation of scores attributed to the suppliance, over-suppliance and misuse of English morphemes in obligatory and non-obligatory contexts thus:

$$\frac{n\ correct\ suppliance\ in\ obligatory\ contexts}{n\ obligatory\ contexts + n\ suppliance\ in\ non\text{-}obligatory\ contexts} \times 100$$

(Pica 1983: 71)

Ellis & Barkhuizen (2005: 80) suggest a different formula for percentage accuracy scoring:

$$\frac{n\ correct\ suppliance\ in\ contexts}{total\ obligatory\ contexts} \times 100$$

Both of these two proposals involve the notion of language required in an obligatory context, a notion which is inappropriate for this study. Because it is possible, as indicated in Chapter 3.4, to avoid the passive in some situations, it cannot always be stated that its use is 'obligatory'. Equally, *counting* small items such as morphemes is considerably simpler than *counting* passives, given that there are a number of morphological variables involved in the latter, not to mention the semantic intricacies that have been described in the course of the preceding chapters. A calculation mechanism based on a combination of Pica (1983) and Ellis & Barkhuizen's (2005) formulae was therefore devised for this part of the analysis. Percentage accuracy of target-like choice is calculated in the following way:

$$\frac{n\ correct\ suppliance\ in\ chosen\ context}{n\ number\ of\ attempts} \times 100$$

In this formula, 'context' refers to the occasion chosen by each individual learner and only correct choices of *werden* or *sein* are *counted*. 'Attempt' is every token produced

by an individual in the specified part of the corpus, whether correctly chosen or not, and represents what we have already termed 'passive intent', i.e. it is the context which the learner feels is obligatory.[1] In other words, accuracy is calculated by identifying the number of correctly chosen auxiliary tokens from among the entire auxiliary token *count*.

For example, in Year 4, A1 made 46 attempts to produce a *werden* passive and 22 attempts to produce a *sein* passive, of which 42 tokens and 5 tokens respectively were deemed to exhibit a correctly chosen AUX. The following calculations were made for A1. The process was repeated for all respondents.

A1 *werden* passives

$$\frac{42}{46} \times 100 = 91.3$$

A1 *sein* passives

$$\frac{5}{22} \times 100 = 22.7$$

The production of each respondent in the A corpus is shown in graph form in Figures 6.1 and 6.2, indicating the total number of *werden* and *sein* passive tokens produced in all four exercises compared to the number of correct choices of AUX made by the learners. Raw scores are used here because each attempt has been evaluated separately; it is not a question of the passives-per-500-words being compared across the corpus but of tracking the learners' use of the auxiliaries in each exercise in order to observe any changing patterns. The number of attempts to produce the passive (what has previously been termed 'passive intent') is mapped against the number of times the learner's choice is semantically correct in the context in which it is used, indicated in the Figures by CC to represent 'correct choice'. Appendix D and the explanation in Chapter 4.8.2.2 clarify a question which may be raised concerning the legitimacy of the researcher's evaluation of 'correct' and 'incorrect': by requesting the input of both an L1 E and an L1 G interrater via a questionnaire, the researcher's judgement was supported. Note that no data are available for A4 Year 1.

---

[1] Ellis & Barkhuizen (2005: 80) point out that their formula does not take account of overuse of an item (in their case, a morpheme) and it is acknowledged that the same is true of the formula adopted here.
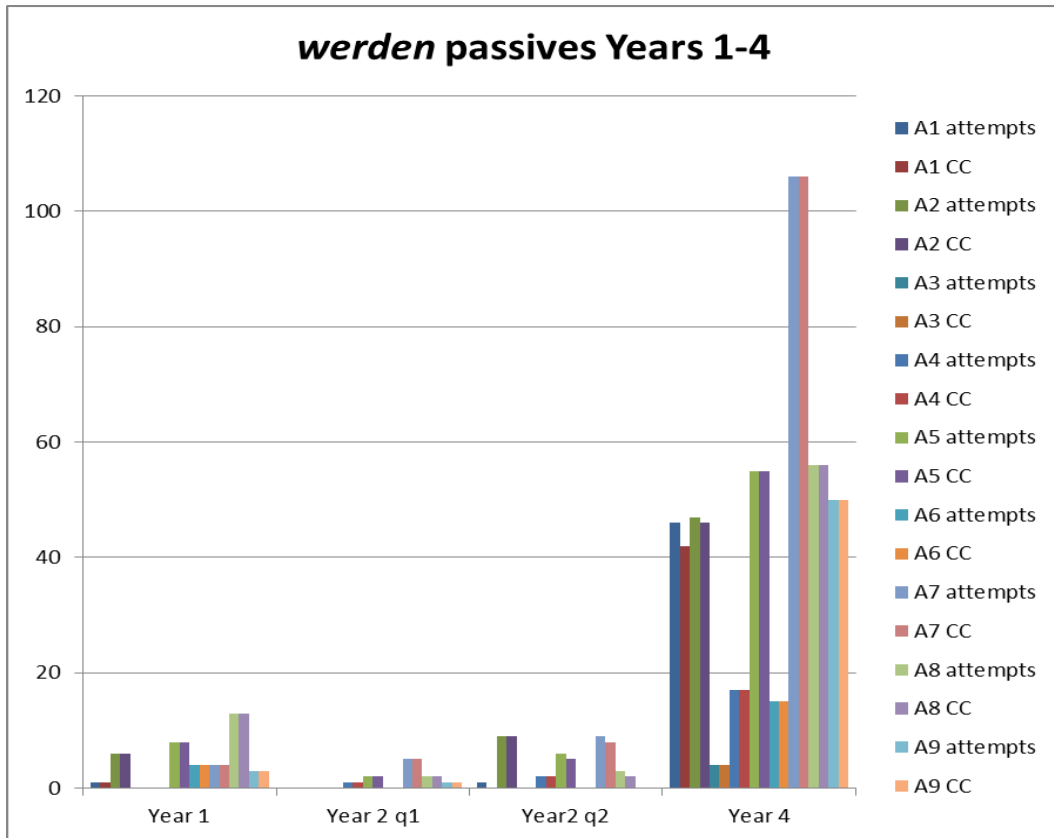
Figure 6.1: Correct choice of auxiliary *werden* in the learners' context (CC), Years 1-4
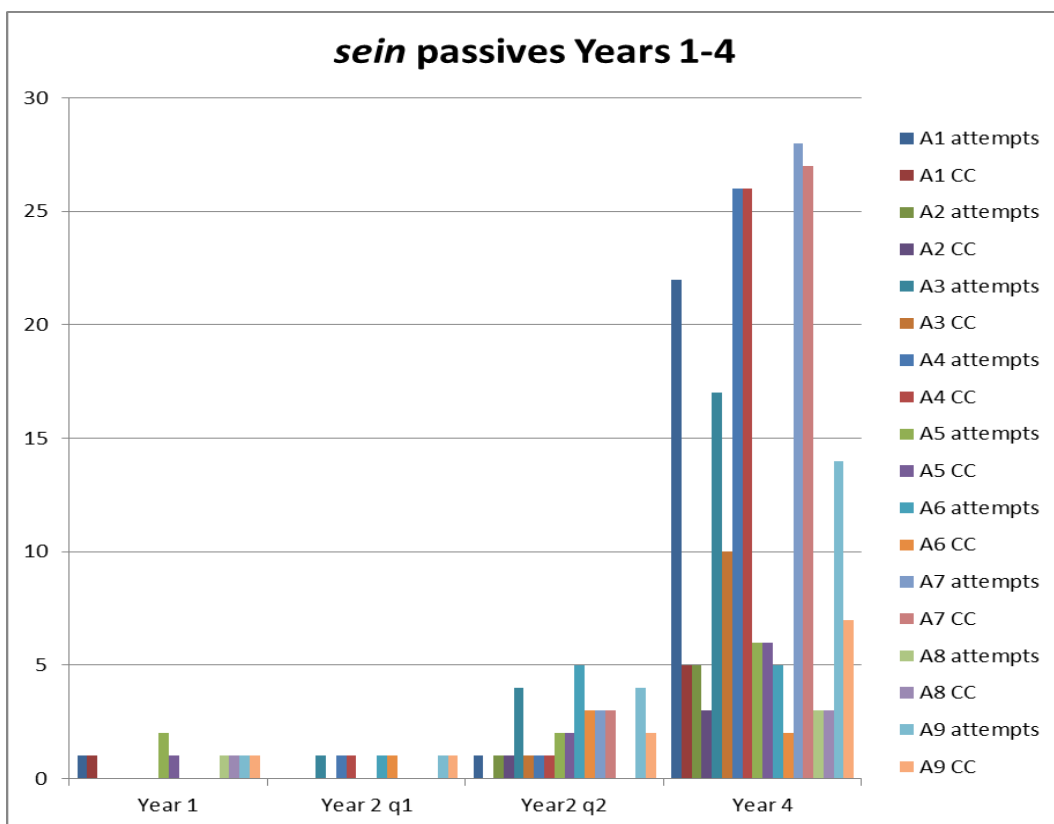
Figure 6.2: Correct choice of auxiliary *sein* in the learners' context (CC), Years 1-4

While accuracy of choice may seem to be greater in the early production of the cohort, it is often the case that no tokens are produced in the sample at that time, as the zero scores show. The *sein* passives of A1, A2, A3, A4, A6 and A9 in Year 4 show the most variation between attempts and correct choice. A9 demonstrates the most steady pattern of use of both types of passive over all four exercises. The lack of passive tokens produced in the first three exercises by A3 will merit further comment in the case study report. The performance of A7 and that of A8 show the greatest degree of correct pairing of form and function, but neither of them achieves a perfect score, with a mismatch in the *sein* passives of A7 (Year 4) and the *werden* passives of A8 (Year 2 question 2).

6.3   The A Corpus Case Studies

The tables in each case study show both the number of occasions on which a passive was attempted by a learner and the result of the calculation to assess percentage accuracy (correctness) of choice (represented by CC). The results of the CC calculations are followed by a discussion of specific items of production from the corpus with respect to each respondent, bearing in mind the intention to assess patterns that may be observable in the learners' production. No claims are made regarding the application of the comments to the rest of the corpus; these remarks serve only to identify points of interest in the learner's current grammar. Because the data sample is small in the first three exercises, comments are often focused on the Year 4 exercise where a more extensive range of use of the passive is produced by the users. Where comments relating to the process of acquisition are made with respect to the Year 4 exercise, a benchmark of 80% correct choice in the learner's context is used. As Palotti points out, 'the representation of acquisition is to some extent arbitrary' (Palotti 2007: 361), and he then provides examples of studies which variously define acquisition in terms of correct usage in '60 per cent (Vainikka and Young Sholten (sic) 1994), 75 per cent (Ellis 1988), 80 per cent (Andersen 1978), or 90 per cent (Dulay and Burt 1974; Bahns 1983) of cases' (2007: 362). Since Vainikka & Young Scholten's low benchmark of 60% was called into question by Sharwood Smith & Truscott (2005) and the results from the latter study are used to support the claims we make in the final section of this chapter, the benchmark of 80%, being roughly the average expectation in the other studies listed by Palotti, seems justified here.

Consideration should also be given, however, to those respondents who make a number of unsuccessful attempts at producing passives and do not achieve the 80% success rate which would in these terms indicate acquisition. In Pienemann's terms, outlined in Chapter 2.5.3.5, a respondent who makes a correct choice five times demonstrates that the construction has emerged.[2] On the other hand, a score of 22.7% CC as illustrated in Table 6.1 would not count as acquired according to the 80% benchmark which is applied in this study.

6.3.1   Student A1 Case Study

Learner	*werden* attempts	CC	% CC	*sein* attempts	CC	% CC
Year 1	1	1	100	1	1	100
Year 2 q1	0	0	0	0	0	0
Year 2 q2	1	0	0	1	0	0
Year 4	46	42	91.3	22	5	22.7

Table 6.1: A1 correct choice of auxiliary in the learner's context (CC)

Table 6.1 indicates that a single token of the *werden* passive was produced by A1 in Year 1. (1) reveals that, although the choice of AUX is correct, the grammatical form of a periphrastic passive has been confused with the use of *man* + an active verb form (see Chapter 3. 4).

(1)   *im Hotel,   wo   man   Abendessen   serviert   wird
       in the hotel   where one/someone dinner   served   becomes/is
       becoming
       in the hotel where dinner is served

While *serviert wird* is typologically and formally correct for the 3pers sing, A1 confuses this with the alternative way of expressing some passives by introducing a second subject i.e. *man,* which is followed by an active verb. In Year 4, a closer look at the quality of the passives produced by the learner indicates that of the 46 *werden* passives attempted, 42 are correct in form and function which represents a score of 91.3%, easily

---

[2] See Chapter 2.5.3.5 for a description of the difference between acquisition and emergence.

151

reaching and surpassing the suggested benchmark of 80% to indicate acquisition. This is not the case for *sein* passives: 22 are attempted but only five are correctly chosen.

6.3.2   Student A2 Case Study

	werden			*sein*		
Learner	attempts	CC	% CC	attempts	CC	% CC
Year 1	6	6	100	0	0	0
Year 2 q1	0	0	0	0	0	0
Year 2 q2	9	7	77.7	1	1	100
Year 4	47	46	97.9	5	3	60

Table 6.2: A2 correct choice of auxiliary in the learner's context (CC)

Few passive tokens appear in years 1 and 2 and no conclusions are drawn from the results apart from the fact that the learner's production of *werden* passives is fairly accurate when it occurs. A2's production of the passive reaches the 80% benchmark for *werden* passives in year 4. The 100% result for *sein* passives in year 2 question 2 is regarded as insignificant because of the low sample size.

6.3.3   Student A3 Case Study

	werden			*sein*		
Learner	attempts	CC	% CC	attempts	CC	% CC
Year 1	0	0	0	0	0	0
Year 2 q1	0	0	0	1	0	0
Year 2 q2	0	0	0	4	1	25
Year 4	4	4	100	17	10	58.8

Table 6.3: A3 correct choice of auxiliary in the learner's context (CC)

The lack of passive typology and tokens throughout this student's performance suggests that this respondent is one of those to whom Han refers, as indicated in Chapter 2.5.3.8, as resisting environmental influences and failing to make progress (Han 2013: 133), even by Year 4. In addition, in the Year 4 exercise, lack of clarity meant that the researcher had in many cases to make a decision concerning whether or not the learner was attempting to produce a passive, as explained earlier in this chapter and in Chapter 4.4.4 and 4.8.2.1 and Chapter 5.1.

(2)   *Sie   waren so   überwältigen, dass…

        *They   were   so   to overwhelm, that…

In (2), by confusing INFIN with PP, the learner makes it difficult for the intended sense to be conveyed. The researcher concluded from examining the context that this token should be regarded as predicative use of an adjectival phrase (= *they were so overwhelmed*) rather than as a dynamic passive (= *they were being overwhelmed*). It should be noted that, as was was pointed out in Chapter 3.4, the construction of *sein* + *zu* + INFIN can easily be confused with other constructions. The increase in *sein* passives in Year 4, particularly in comparison with the low sample score for *werden* passives is suggestive of overuse.

6.3.4   Student A4 Case Study

	werden			*sein*		
Learner	**attempts**	**CC**	**% CC**	**attempts**	**CC**	**% CC**
Year 1	N/A	N/A	N/A	N/A	N/A	N/A
Year 2 q1	1	1	100	1	0	0
Year 2 q2	2	2	100	1	1	100
Year 4	17	17	100	26	12	46.1

Table 6.4: A4 correct choice of auxiliary in the learner's context (CC)
N/A = no data available

Confining comments to the Year 4 exercise because of the low sample sizes in the Year 2 exercises, A4's low accuracy score for *sein* passives suggests that this type is not acquired, although in Pienemann's terms it could be considered as having emerged. *Werden* passives, on the other hand, are stable and acquired by year 4.

6.3.5   Student A5 Case Study

	werden			*sein*		
Learner	**attempts**	**CC**	**% CC**	**attempts**	**CC**	**% CC**
Year 1	8	8	100	2	1	50
Year 2 q1	2	2	100	0	0	0
Year 2 q2	6	5	83.3	2	2	100
Year 4	55	55	100	6	6	100

Table 6.5: A5 correct choice of auxiliary in the learner's context (CC)

A5 achieves and surpasses the 80% benchmark in all attempts at the passive except for Year 1 *sein*, even under the pressure of restricted time in year 2, although the sample size in that exercise is low. From these results the conclusion is drawn that control of the form/function relationship has been stable since the early stage for *werden* passives and has become so in the case of *sein* passives. Limited production of the construction in year 4 could simply indicate the learner's preference for other means of expression and these options are not explored here. It could be that this learner is cautious when using passives, being aware of the possible ambiguities and L1 interference associated with *sein* as explained in Chapter 2.2.2.2, and seeking to avoid errors by selecting *werden,* the meaning of which (= *become*) is closer to the 'get' passive, described in Chapter 3.2.2.3.

6.3.6   Student A6 Case Study

	werden			*sein*		
Learner	**attempts**	**CC**	**% CC**	**attempts**	**CC**	**% CC**
Year 1	4	4	100	0	0	0
Year 2 q1	0	0	0	1	1	100
Year 2 q2	0	0	0	5	3	60
Year 4	15	15	100	5	2	40

Table 6.6: A6 correct choice of auxiliary in the learner's context (CC)

In the case of A6, when *werden* passives are produced, they are stable. It is interesting that this learner, like A3, produces no *werden* passives in the year 2 question 2 exercise where the passive in the title may have acted as a prime.

6.3.7 Student A7 Case Study

	werden			sein		
Learner	attempts	CC	% CC	attempts	CC	% CC
Year 1	4	4	100	0	0	0
Year 2 q1	5	5	100	0	0	0
Year 2 q2	9	8	88.8	3	3	100
Year 4	106	106	100	28	27	96.4

Table 6.7: A7 correct choice of auxiliary in the learner's context (CC)

The production of A7 was examined in the preliminary study as a *comparison* to that of A8. It was noted in Chapter 5.3.1.2 that there were similarities between the incidence of *werden* passives of A7 and A2 in Year 2 examination question 2 and in the ratio of *werden* passives to *sein* passives in the same exercise between A7 and A5, but that otherwise no parallels between A7 and the rest of the A respondents could be seen. It should be mentioned, however, that similarities in respect of the particular analyses which have been carried out were few in the written production of the L1 E learners; the evidence provided thus far supports the contention that although some individuals exhibit control of the form/function relationship in the German passive from the early stage of this study, the extent to which the construction can be said to have been acquired varies across the learner cohort.

6.3.8 Student A8 Case Study

	werden			sein		
Learner	attempts	CC	% CC	attempts	CC	% CC
Year 1	13	13	100	1	1	100
Year 2 q1	2	2	100	0	0	0
Year 2 q2	3	2	66.6	0	0	0
Year 4	56	56	100	3	3	100

Table 6.8: A8 correct choice of auxiliary in the learner's context (CC)

Student A8, it should be remembered, was the L1 E subject in the preliminary study which was carried out in the first phase of this project. In the preliminary study analysis it was noted that A8 produced more instances of the *wird* passive-per-500-words in her

shortest (Year 1) text than in either of the other two exercises which were assessed in that section of the study. This finding was borne out when passives-per-500-words were calculated during the main analysis, as Table 5.18 indicates. We conclude that A8's production of the *werden* passive in year 1 was indicative of reliance on formulaic language, since four tokens of the nine *wird* passives that were counted were the same: *wird* (AUX) *verurteilt* (PP) (= is being convicted). Formulaic language is here defined as in Chapter 2.5.3.6 as 'a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar' (Wray 2002: 9). German sentence structure requires AUX and PP to be separated, thus discontinuous in Wray's terms, and repeated use of the same grammatical verb form (*wird verurteilt*) is presumably less time consuming in terms of processing than retrieving a grammar rule from memory and producing an alternative construction. Nevertheless, we conclude from the high accuracy scores achieved by A8 that, similarly to A5, control of the form/function relationship has been stable since the early stage. We note that the incidence of *sein* passives is low but suggest that this is a matter of learner choice.

6.3.9 Student A9 Case Study

	werden			*sein*		
Learner	**attempts**	**CC**	**% CC**	**attempts**	**CC**	**% CC**
Year 1	3	3	100	1	1	100
Year 2 q1	1	1	100	1	1	100
Year 2 q2	0	0	0	4	2	50
Year 4	50	50	100	14	7	50

Table 6.9: A9 correct choice of auxiliary in the learner's context (CC)

It was noted in Chapter 5.3.2, commenting on the results shown in Table 5.18, that A9's production of both *werden* passives and *sein* passives per-500-words declined from year 1 to year 4. In addition, it can now be stated that accuracy of choice in the case of *sein* passives is only half that of *werden* passives in year 4 where the sample size is large enough to merit a comment.

6.4   The B Corpus

Table 6.10 indicates that, almost across the board, the B respondents demonstrate complete accuracy of choice of AUX. It should be noted, however, that even native speakers make mistakes.[3] B4 and A7 achieve less than perfect scores. The discrepancy in the percentage totals for these two respondents is addressed in Chapter 8.

Respondent	Word count	*werden passives* attempts	CC	% CC	*sein passives* attempts	CC	% CC
B1	9183	119	119	100	27	27	100
B2	11561	115	115	100	33	33	100
B3	11414	164	164	100	24	24	100
B4	1963	57	57	100	4	3	75
A7 year 4	6825	106	106	100	28	27	96.4

Table 6.10: The B corpus correct choice of auxiliary in the user's context (CC)
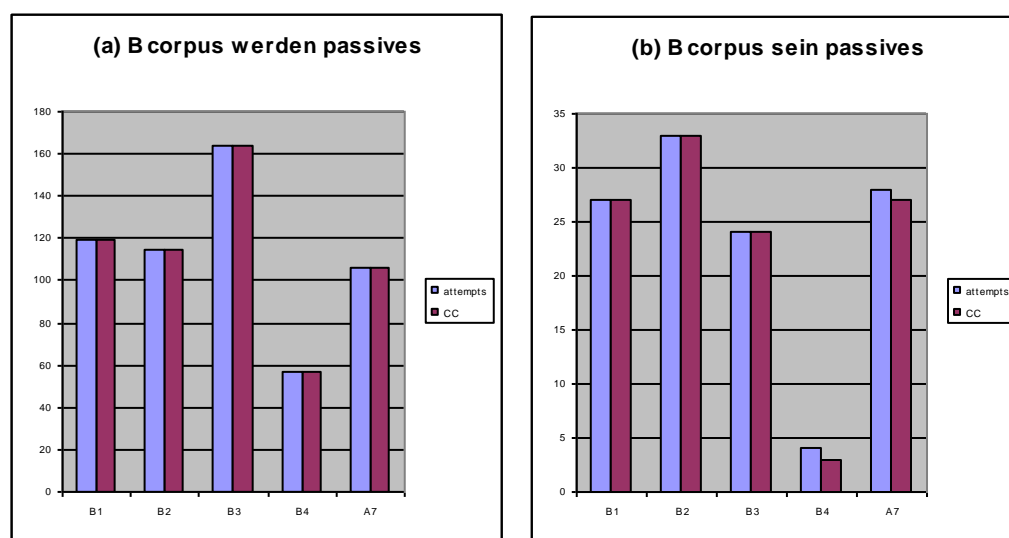
Figure 6.3 (a) and (b): The B corpus correct choice of auxiliary in the user's context (CC)

As in Chapter 5.4, A7's production in Year 4 is benchmarked against a similar (dissertation) exercise produced by the B respondents.   The comparison between the accuracy of target-like choice in respect of AUX in the B corpus and the A corpus is shown in diagrammatic form in Figures 6.3 and 6.4. A7 is represented in both figures.

---

[3] The identification of errors in the production of B4 and A7 was confirmed by the L1 G interrater.

It can be seen that A7's production is more similar to that of the B respondents in the L1 G corpus than it is to that of the L1 E learners.
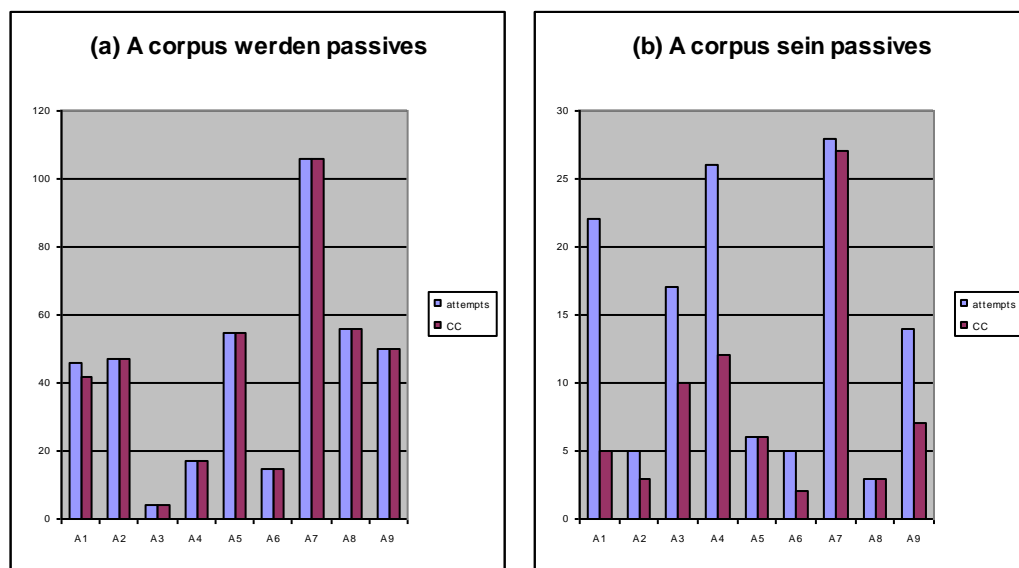
Figure 6.4 (a) and (b): The A corpus correct choice of auxiliary in the learner's context (CC)

6.5   Quantitative Analysis: C Corpus

The sample of input data in the C corpus is small but gives some indication of the relative variety of FL texts, either collected or produced in written format by the tutor, which would commonly be distributed to students in preparation for their own writing, as explained in Chapter 4.6.1.3.

Item		
1	*wird*	14
3	*werden*	21
5	*wurde*	8
7	*wurden*	4
9	*worden*	1
	TOTAL	48

Table 6.11: *werden* passives, C corpus

Item		
2	*ist*	3
4	*sind*	10
6	*war*	0
8	*waren*	0
	TOTAL	13

Table 6.12: *sein* **passives, C corpus**

Total number of words in the C corpus: 3158

Present tenses feature more often than past tenses in the Year 2 input shown in the C corpus, with the 3pers pl occurring most frequently. In the students' production in the Year 2 examination question 2, as indicated in Chapter 5.3.1.2, both singular and plural forms appear almost equally, suggesting that production may have been influenced by the input, some of which was tailored to a similar essay question.

It is worth noting that not all the passive items on which we have focused in this analysis are included in the input (items 6 and 8 are missing) and this could account in part for non-nativelike use on the part of the A cohort. The fact that Duden's ratio of 5 *werden:* 2 *sein* is not demonstrated in this small sample is an indication that these texts are not necessarily a representative sample of contemporary German texts.

In the case of the Year 4 dissertation which was prepared during the A cohort students' study abroad, no evidence of input is available. It is likely that daily exposure to varying types of input (e.g. written, spoken, explicit grammar instruction) will have influenced the students' writing, but in this study we can only speculate that this may be the case.

6.6   Alternatives to the Periphrastic Passive

Duden lists a number of caveats in its description of the German passive, as we have seen in Chapter 3.4. Among the alternatives to the periphrastic passive which are mentioned are two types of infinitive constructions, those combined with the reflexive verb *sich lassen* (= to allow something to be done) and those combining *sein + zu +*

INFIN (= something is to be done), the latter used as an alternative to a *werden* passive. We shall refer now to the production in the corpus of these two constructions, the first of which appears rarely, the second, more frequently, especially in the B corpus.

In the formation of the passive, the verb *sich lassen* has only 2 morphological forms: the 3pers sing and 3pers pl. Its sense is completed by the addition of a verb final infinitive.[4] The present and simple past singular and plural forms were entered into the WordSmith 4 (Scott 2004) concordancer. Usage was minimal and only the results for those respondents who produced a *sich lassen* construction are recorded in Tables 6.13, 6.14 and 6.15. Five respondents produced a *sich lassen* form in year 1; one of them (A8) produced the same 3pers sing form in examination question 1. No items were found either in the Year 2 examination question 2 or in the Year 4 exercise. It should be noted that these *sich lassen* constructions are all produced correctly in morphological and syntactic form. In the case of the 3pers pl form (*lassen sich*) used by four respondents in year 1, this form is the same as the infinitive and repeating it requires little in terms of processing capacity. It is possible that this construction is acquired by the learners as a formulaic sequence: a simple formula such as *es lässt sich* with the addition of a final position infinitive is probably relatively straightforward to memorise.

Item	A1	A2	A3	A8	A9	Total
lässt sich	0	0	1	1	0	2
lassen sich	1	1	1	0	1	4
TOTAL	1	1	2	1	1	6

Table 6.13: *sich lassen* concordances, Year 1, A corpus

Item	A8
lässt sich	1
TOTAL	1

Table 6.14: *sich lassen* concordances, Year 2, exam question 1, A corpus

---

[4] An example of this construction appears in Chapter 3.4.

Item	B1	B3	B4	A7	Total
lässt sich	3	4	1	7	**15**
lassen sich	5	2	1	1	**9**
ließen sich	1	0	0	0	**1**
TOTAL	9	6	2	8	**25**

Table 6.15: *sich lassen* concordances, B corpus[5]

Looking at Tables 6.13 and 6.14, we notice once again that A8 features in the positive result. She is clearly aware of alternatives to the periphrastic passive from year 1 and is also capable of producing one token under time constraints. She is not the only respondent to use this construction in Year 1, yet the absence of any further tokens in the production of the other L1 E respondents is suggestive of formulaic use in the early stages which then disappears from the output across the corpus. In Year 4 in the A corpus, tokens of this construction appear only in the case of A7, the NS of German. This leads us to suggest that the *sich lassen* construction occurs in nativelike language which even advanced learners with L1 E do not replicate. This suggestion is borne out by the incidence, although small, of tokens of the construction in the production of the B respondents, with whose written production A7's year 4 exercise is compared, as shown in Table 6.15.

Turning to the infinitive constructions with *sein + zu*, Tables 6.16, 6.17 and 6.18 indicate the frequency with which these occur in the corpora. They are neither *werden* passives nor *sein* passives although they function as semantic passives. The grammatical subject position may be occupied by a clause with the specific agent unknown, as (3) from A7's year 4 exercise demonstrates. English modal verbs usually offer a useful translation, as in (3), (4) and (5) below.

(3)　　Wenn es　　　nun so einfach ist,　　die Insel zu verändern,　　　ist

　　　　deshalb　　　anzuzweifeln.

　　　　If it　　　　so easy　　　is,　　the island to alter,　　　　is

　　　　therefore　　to doubt.

　　　　It is doubtful whether the island can be altered so easily.

---

[5] Both the forms with –ss and –ß spelling were entered into the concordancer. Only the positive samples were recorded.

A1	A2	A3	A4	A5	A6	A8	A9	Total
1	1	1	N/A	0	0	1	0	4

Table 6.16: *sein* + *zu* + **INFIN, Year 1, A corpus**

A1	A2	A3	A4	A5	A6	A8	A9	Total
0	0	1	0	0	0	0	0	1

Table 6.17: *sein* + *zu* + **INFIN, Year 2, exam question 1, A corpus**

B1	B2	B3	B4	A7	Total
20	46	10	3	9	88

Table 6.18: *sein* + *zu* + **INFIN, B corpus**

We note that the A cohort displays very few instances of these alternatives to the passive and that these are almost entirely limited to production in year 1. No items were found either in the Year 2 examination question 2 or in the Year 4 exercise except in the production of A7, shown benchmarked against the 4 members of the B cohort in Table 6.18. A closer examination of the items suggests that these indicate formulaic use: the Year 1 exercise required a tourist brochure to be written and (4) (from A1) and (5) (from A2) may well have occurred in the German preparatory materials which the students examined.[6] This purely speculative suggestion is based on pedagogical experience. A statement with more robust foundations could have been produced if the input materials had been available.

(4)   Der ganze Morgen   ist   hier   zu verbringen.
        The whole morning   is   here   to spend.
        You can spend the whole morning here.

(5)   Die „Heimatmuseen"   sind   zu besichtigen.
        The local history museums   are   to visit.
        Visiting the local history museums is a 'must'.

In stark contrast to the L1 E respondents, the members of the L1 G cohort opt far more often for *sein* + *zu* + INFIN in order to render a passive. No calculation of passives-per-500-words was carried out in this part of the analysis, as alternatives to the passive are not the primary focus of the study.

---

[6] Formulaic language is defined in Chapter 2.5.3.6 and further comments were made in Section 6.3.8.

6.7   Concluding Remarks on Chapter 6

Chapter 6 commenced with an explanation of the implementation of a formula-based data analysis which was applied to the A and B corpora. The case study format in which the results of the calculations were presented enabled a *comparison* to be made between the production of the passive not only by individual learners over time but also between learners with the same mother tongue (English) and between NNSs and NSs of German, representing the contrastive interlanguage analysis described in Chapter 2.6.2. In the remarks which conclude this chapter, it will be possible to draw on the case study commentaries and observe any interesting patterns which emerge from this analysis.

As suggested in Section 6.3, because of the small sample size it is not possible to comment on whether or not the passive has been acquired in the year 1 and year 2 exercises, since some of the results indicate only one or two tokens produced. Using the accuracy measure described in Chapter 2.5.3.5 and in Section 6.3 to analyse the year 4 exercise, however, it is possible to say that all learners have acquired the *werden* passive, with the exception of A3 whose sample size (4 tokens) is too small even to indicate emergence despite the fact that all 4 attempts are correctly chosen in the learner's context. In respect of *sein* passives, none of the L1 E cohort, with the possible exception of A5 with 6 CC tokens, has reached the point of acquisition. It is, however, important to note that A7 with her special status as the respondent who studied along with the L1 E cohort but is a NS of German, has a perfect score of 100% accuracy of CC with 106 items of *werden* passives and 96.4% accuracy with *sein* passives. It is interesting that she, too, like the rest of the A cohort, has a lesser score with *sein* passives than with *werden* passives. If any CC score of more than 5 signifies emergence, as has been suggested in Section 6.3, prior to year 4 this would only apply to A5 and A7 (*werden)*, since A8's score of 13 CC (*werden*) in year 1 could be attributed to formulaic use of the construction for the reasons advanced in Section 6.3.8.

The results certainly indicate that there is considerable variation in the production of the passive between learners. As has been pointed out, evidence of entrenched errors is seen in A3 at a low level and A9 at a higher level, while in contrast A5 and A8 demonstrate constant, accurate control of the form/function relationship. In addition, A6's use of the *sein* passive which appears to have regressed in year 4 would appear to

support Sharwood Smith & Truscott's (2010) contention that acquisition occurs along a continuum which may sometimes involve regression as well as growth.

The almost 100% accuracy of choice of *werden* passives by the L1 E learners in Year 4, compared to significantly less control in the case of *sein* passives, is shown in Figure 6.4. The suggestion is that acquisition of the semantics of the *sein* passive lags behind that of the *werden* passive in the majority of cases in the A corpus even if the passive syntax is acquired i.e. learners can correctly produce a passive of a sort but it is not the correct sort in their chosen context. This would seem to support recent claims that problems for learners occur at the interface between syntax and semantics (e.g. Slabakova 2013; Sorace & Filiaci 2006, as discussed in Chapter 2.5.3.7). The accuracy of choice and the use of other, alternative constructions to express the passive displayed in the language of the B corpus suggests that NSs have more variables available in their mental grammars and that they do have the capacity to process more information simultaneously than NNSs, as suggested by VanPatten & Williams (2007: 116).[7]

While it was not possible to consider the relevance of FL input on the language development of the A cohort during their year abroad, a small sample of input during their second year studies, analysed in Section 6.5, was included. The size of the sample and the relative lack of evidence extracted from it do not provide conclusive evidence of its role in the acquisition of the passive.

In the next chapter, a closer examination of the written production of passives by the L1 E learners of German in year 4 is reported. This is the final phase of the CIA, in which learner language and NS language are *compared*.

---

[7] See Chapter 2.3.4.1.

CHAPTER 7 – A COMPARISON OF NON-NATIVE-SPEAKER PASSIVES WITH NATIVE-SPEAKER EQUIVALENTS

7.1   Introduction

The point has been made throughout the thesis that the token *count* of passives in the year 1 and 2 exercises is too small for any conclusions to be extrapolated from the data in order to make valid generalisations relating to the acquisition of the passive by all learners with L1 E. The general aims in this study, first stated in Chapter 1.7, are to examine patterns of use of the German passive by adult learners with L1 E and to suggest reasons for those patterns, basing the comments on written production collected in a learner corpus which in due course may be used by other researchers who require corpus evidence in their studies. The corpus data provide the empirical sample for the observations which are made - corpus evidence is not intended to contribute to a theoretical standpoint, since no specific theory is being tested. This study is concerned with the correct pragmatic use of the passive only in so far as it observes the extent to which learners with L1 E differentiate accurately between actional and statal passives. Its focus, as stated in Chapter 1. 3 and reiterated in Chapter 6.1, is on the use of the syntax of the German passive by a group of learners whose L1 has fewer ways of expressing it. Additionally, as was mentioned in Chapter 6.2, it is not possible to state that use of a passive is obligatory in a particular context in German, since there are a number of options available as alternatives. The use of three of these alternatives is explored in Section 7.5.

Some observations have been made, however, on the production of passives by individual respondents in the A cohort following both the cross-sectional and the longitudinal analyses in Chapters 5 and 6, and a brief survey of NS German passives retrieved from the B corpus was included in Chapter 6.4. The longer text length of the Year 4 exercise written by the A cohort offers a greater number of examples of *werden* and *sein* passives than the exercises from years 1 and 2 and, using the CIA methodological approach of comparing the IL of NNSs with the production of NSs of the same language, an approach which was introduced in Chapter 2.6.2, the passives produced by the A cohort will be *compared* in this chapter to the data collected in the B corpus, commenting on patterns of passive use, or the lack of them, observed in both the

A and the B corpora. In this survey, the production of A7, the NS of German who studied with the A cohort, is included in the B cohort along with the other L1 G respondents. Following the transfer of the A7 data to the B corpus, the size of the A and B corpora is almost equal, with the B corpus being slightly larger (a total of 40,946 words) than the A corpus (40,075 words), offering a suitable platform for the purposes of *comparing* the two datasets according to CIA methods.[1]

In the course of the thesis thus far, suggestions have been made regarding what might be hypothesised to appear in terms of passives in the production of L1 E learners and reference was made in Chapter 3.5 to the ratio of *werden* to *sein* passives found in Duden's (1998) contemporary German texts; the commentary in this chapter will refer to those points and, in *comparing* the passives in the two corpora, consider the differences between the written production of individual respondents in the A and B corpora and suggest some reasons for those differences. Section 7.4 considers the notion of ultimate attainment, and the question of how differences between the IL of the learners and the production of the NSs might be accounted for is considered throughout the chapter.

In the studies of the acquisition of L1 German passives reviewed in Chapter 2.2.2.2, three of the four, Mills (1985), Bryan (1995) and Abbot-Smith (2003) provided evidence to support the view that child learners of German produce statal passives (*sein* passives) before actional passives (*werden* passives).[2] From this it might be supposed that if L2 learning follows the same sequence as L1 learning, L2 learners of German also acquire statal passives before actional passives, in particular because in English the passive is based on 'be' (*sein*) and also because adult learners of German L2 are already equipped with the conceptual knowledge which enables them to distinguish between passive and active, whereas children have to understand the pragmatic or semantic sense of the concept at the same time as mastering the grammatical elements of the construction.

However, as was suggested in Chapter 2.4, the claim here is that the reverse is the case. Taking the meaning of 'acquired' as defined in Chapter 2.5.3.5, that is, users exhibit

---

[1] The word counts in the A and B corpora were given in Chapter 4.6.1 and in A7's Year 4 exercise in Table 5.15.
[2] The fourth study reviewed (Eisenbeiß 1993) indicates that the two passive types are acquired concurrently in L1.

80% accuracy of both form and function in their chosen context, the production of the two periphrastic passives and of the alternative constructions *sein + zu +* INFIN, *sich lassen +* INFIN and *man* with an active verb form is subjected to closer scrutiny in this chapter, building on the numerical *count* of CC and the mini case studies in Chapter 6. The attainment of individual members of the A cohort exhibited in their use of periphrastic passives and the three alternatives to the passive is *compared* with the production of individuals in the B cohort in order to comment on, and offer possible reasons for, observable differences between the two, thus adding a further dimension to the response to research question 4, which is:

What similarities or differences can be observed between the use of the passive by NSs and NNSs in the respective sub-corpora?

Some examples of the periphrastic *werden* and *sein* passives which are produced by the A and B cohorts will be considered in Section 7.2, in connection with verb types according to Duden's categorization, described in Chapter 3.3.1.1, followed by agent 'by' phrases, described in Chapter 3.3.3 and finally alternatives to the passive which were the subject of Chapter 3.4.

7.2   Use of the Passive

Figure 7.1 indicates the total number of *werden* and *sein* passives per 500 words (what has previously been classified in this study as 'passive intent' in Chapter 4.4.4 and operationalised in Chapter 5.1) exhibited in the NNS A corpus *compared* with the number of tokens of these passive types extracted from the NS B corpus. The figure is compiled from the data shown in Tables 5.18 and 5.21 and does not show pragmatically accurate use of the passive by either the A or the B cohort. This has been discussed in Chapter 6, Sections 3 and 4, whereas here each attempt to produce a *werden* or *sein* passive is tabulated, whether morphologically accurate or not. Morphological accuracy is considered in Section 7.2.3.

If the NSs are taken as the benchmark, it is possible to suggest from observing Figure 7.1 that the learners with L1 E underuse the *werden* passive and overuse the *sein* passive. This suggestion accords with some observations which were made in the context of the cross-sectional and longitudinal *counts* in Chapter 5. In addition, it

should be remembered that NS use of the passive was shown in Chapter 6.4 to be almost 100% accurate in the pragmatic decision of whether to produce a statal or an actional passive, whereas the analysis of NNS use of the two types in Chapter 6.3 led to the conclusion that their pragmatic use of *sein* passives is only half as accurate as that of *werden* passives in year 4 where the sample size is large enough to merit a comment.
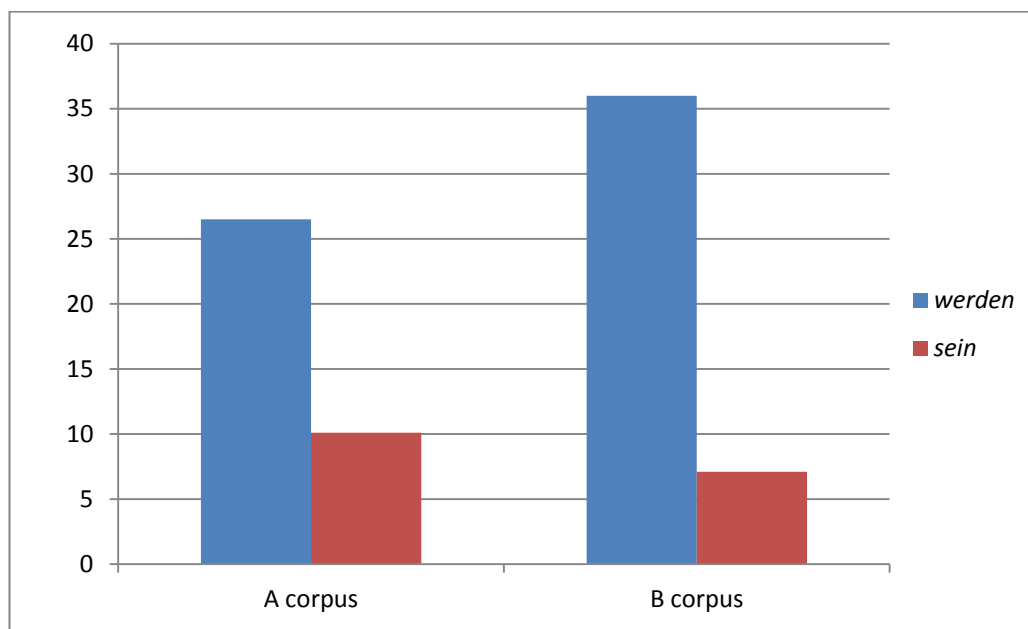
Figure 7.1: Total *werden* and *sein* passives-per-500-words; A and B corpora

It is, however, worthy of note that the preponderance of *werden* passives over *sein* passives is similar in both of these equal-sized corpora, with the B corpus being closer to Duden's (1998) ratio of 5 *werden* : 2 *sein* than the A corpus, suggesting that the respondents in the current LINCS corpus as a whole are not demonstrating NS-like usage, even though one or two individuals (e.g. A5) have 'acquired' the pragmatic purpose of the two passives according to the criteria applied in this study. As pointed out in the example of respondent A5 in Chapter 6.3.5, learners may hesitate to use *sein* because of its association with L1 E 'be' and its use as a copula verb, for example, and therefore opt for *werden* because they perceive an approximation between 'become' (*werden*) and a 'get' passive.

In discursive academic dissertations such as those which provide the data for the current analysis, use of the passive is expected as a way of identifying information which has been gathered to support an argument and to evaluate it without expressing personal opinions. There is no suggestion that a certain number of passives is expected in these

exercises, as previously indicated in Chapter 5.3.1. However, with respect to passive morphology, L1 E learners at intermediate and higher levels, such as those whose production is recorded in the A corpus, would be expected to have sufficient grammatical knowledge after three-and-a-half years of undergraduate study to enable them at least to produce correct grammatical forms even if they have not yet mastered the semantic/pragmatic function. It would appear that the distinction between syntax and pragmatics is problematic for some learners, as will be seen in Section 7.2.2.

There is evidence in the corpus that allows the speculation that the L1 E learners in the A cohort use the passive as a means of avoiding the first person *ich* form since personal opinions are discouraged in academic writing unless specifically requested.

(1)   Bevor   die Dissertation   angefangen   wird

         Before   the dissertation   begun   becomes/is becoming

         Before the dissertation is begun/Before beginning the dissertation

         (Student A8)

There is also evidence that passives are used where no indication of agent is given or where none is required, as described in Chapter 3.2.2.1.

(2)   da die Unterschiede zwischen Ost und West and   die   nächste

         Generation   weitergegeben   werden

         as the differences between east and west   to (L1 error)   the   next

         generation   passed on   become/are becoming

         as the differences between east and west are passed on to the next generation

         (Student A3)

7.2.1 Verb Typology

As indicated in Chapter 3.3.1.1, Duden's verb types B and C are rare in contemporary German texts and it can be assumed that they will also be rare in learner production, not simply because they occur infrequently in the input but because they require additional

processing capacity from the learner. Indeed there is only one example of a Duden Type B verb in the entire A corpus (3).

(3)   Studienabsolventen   wurden   dieselben   Fragen
       auch   gestellt
       To graduates   became/were becoming   the   same   questions
       also   put

       Graduates were also asked the same questions
       (Student A5)

Conversely, respondent B3 alone produces at least two passives with Type B verbs, seen in (4) and (5).

(4)   Dazu   werden   dem Teilnehmer   neun
       verschiedene Aussagen   präsentiert
       In addition   become/are becoming   to the participant (DAT)   nine
       different answers   presented
       In addition, the participant is offered nine different optional answers
       (Student B3)

(5)   Mittels   der Profile   wird   den Nutzern
       ein schneller Selektionsprozess […] ermöglicht
       With the help of   the profile   becomes/is becoming to the users (DAT)
       a faster selection process [   ]   made possible

       Producing a profile speeds up the selection process for the users
       (Student B3)

The use of the passive in (4) and (5) compared to the single occurrence of a Duden type B verb in the entire A corpus suggests that NSs of German are more able to cope with what Lardière calls 'computational complexity' than learners whose resources in terms of processing are not as great (Lardière 2003: 689).

7.2.2 Passive Morphology

The passive morphology required for the construction of the German periphrastic passive as described in Chapter 3.3.1 is a PP. Learners with L1 E make few errors with PP morphology, whether produced in conjunction with *werden* or with *sein* or with regular (6) or irregular (7) verbs. As has been indicated previously in this thesis (e.g. Chapter 1.3), the prime focus of the study is the production of German passive syntax by learners with L1 E. Comments on the accuracy of the passive morphology produced by the learners are included at this juncture only insofar as it constitutes half of the necessary syntax required to produce a periphrastic passive. (6) – (9) are error-free.

(6)   das   jährlich   […]   veröffentlicht   wird
        which   annually   […]   publicised   becomes/is becoming

        which is published annually
        (Student A9)

(7)   wird   die Muttersprache   noch   zu Hause   gesprochen
        becomes/is becoming the mother tongue   still   at home   spoken

        the mother tongue is still spoken at home
        (Student A9)

(8)   Zwei von den SPD Wählern   sind   von der sozialen Komponente
        dieser Partei   überzeugt
        Two of the SPD voters   are   by the social contents
        of this political party   convinced

        Two of the SPD voters are convinced by the social policies of this political party
        (Student A8)

(9) Die Mehrheit der Länder sind von CDU/CSU regiert

The majority of the states are by CDU/CSU governed

The majority of the German states are governed by the CDU/CSU
(Student A1)

(10) *Es ist vielleicht erwartete

*It is perhaps he/she/it expected (3pers sing preterite)

*It is perhaps he/she/it expected
(Student A1)

On the other hand, not only is there an error on PP in (10), a *sein* passive has been used instead of a *werden* passive by A1. We may deduce that the two errors in this four-word sequence are caused by the learner attempting to produce a passive and being influenced by L1 English 'is' *(ist)*, and being unable simultaneously to produce the requisite passive morphology on PP (*erwartet*[3]). Such a deduction can only be confirmed by questioning the learner. Since this particular learner produced 42 CC out of 46 attempts at *werden* passives in the qualitative analysis reported in Chapter 6.3.1, in contrast to only five CC of 22 attempts at *sein* passives, we may reach the conclusion that acquisition of the *sein* passive in her case has not yet been achieved.

Similarly, in (11), a *werden* passive has been selected instead of a *sein* passive by A2.

(11) *Die meisten Stadtteile werden von
waldbedeckten Hängen umgeben
*The most parts of the town become/are becoming by
tree-covered slopes surrounded

Most parts of the town are being surrounded by tree-covered slopes
(student A2)

The meaning of the student version with *werden* is that the slopes are actively moving to surround the town, a meaning which is presumably not intended.

---

[3] *Erwarten* (= to expect) is a regular verb but the prefx *ge-* on PP is omitted on verbs commencing *er-*.

However, the same student also produces correctly chosen *werden* passives (12), indicating that either the learner has made an inadvertent slip or mistake in (11) as defined in Chapter 2.5.2.5 (Corder 1981: 10) or that the choice of the native-like auxiliary remains in her grammar as an option which has not been fully resolved, as described by Sorace (2000). This may also be considered as an instance of fossilization if Han's (2012) observation is taken into account. Han remarks that '[f]ossilization, after all, is not zero learning but incomplete learning' (2012: 481), suggesting that A2's use of the *werden* passive is still in the process of change. The inclusion of an additional methodological treatment such as a stimulated recall exercise (e.g. as explained in Gass 2013: 50) would be necessary in order to confirm or disconfirm these suggestions. Such a treatment would enhance future research into the acquisition of the German passive by L1 E learners.

(12) Sofern es   nicht zuerst   von der Verwaltung   genehmigt
      wird
      Unless it   not first   by the governing body   approved
      becomes/is becoming

      Unless it is first approved by the governing body
      (Student A2)

There are only two examples of an incorrect choice of auxiliary in the B corpus (A7 and B4; see Figure 6.3 (b). These may also be regarded as slips.

7.2.3   Morphology of the Auxiliary

Errors are more likely in the learners' production of AUX morphology. That is to say, there are instances in the A corpus of nativelike choice of passive type (*werden* or *sein*) but non-nativelike grammatical form on AUX (13). These are few, however, and without further checks with the learner could just as easily be accounted for as slips in proof-reading as grammatical errors occasioned by the learner focussing on the production of a complex construction.

(13) *Studiengebühren (pl)   wird (3pers sing)   nicht jetzt   eingeführt
      *Tuition fees   is becoming/becomes not   now   introduced

*Tuition fees is not being introduced now

(Student A6)

7.2.4 Formulaic Language

It was suggested in Chapter 5.3.1.1 that expressions learnt as formulaic sequences (Wray 2002: 9) might be present in the learners' IL because they are more readily produced from memory in a needs-only analysis. For example, in the learner's chosen context, a passive is required, and some individuals may rely on an expression which can be reproduced more quickly than one which requires more processing time. Such easily-remembered formulae, like the frequent use of *wird verurteilt* (= is /is becoming judged) by A8 reported in the preliminary study in Chapter 5.2, occur in the learners' IL as a precursor to more advanced structures. A further illustration is to be found in the comparison of the year 4 NNS data which is discussed in this chapter in relation to the NS corpus. While A8 repeatedly produces a straightforward SVO order, albeit a passive, '*es wird erwähnt, daß*..' (it is mentioned that..), this particular expression does not appear in the B corpus at all. Instead, the verb *erwähnen* occurs twice in the B corpus in a different and more complex grammatical context (14).

(14) Wie   bereits   oben   erwähnt
      As   already   above mentioned

      As already mentioned above
      (Student A7)

and the item is used in its 3pers sing present tense form (15) and as a PP in a periphrastic passive (16) by B2.

(15) Als Regelbeispiel   für eine   erhebliche   Behinderung des
      wirksamen   Wettbewerbs erwähnt   Art.2   X
      As example   for a   significant   barrier of the
      efficient   competition mentions   article 2   X

      Article 2 mentions X as an example of a significant barrier to efficient competition

174

(16)   Das Kriterium (…)   war   vorher   zwar   erwähnt
        The criterion (…)   was   previously   indeed   mentioned

        The criterion (…) was of course mentioned previously
        (Student B2)

In the discussion of the learners' choice of auxiliary in Chapter 6, Figure 6.4 indicated that learners with L1 E match the form and function of *werden* passives sooner than *sein* passives and interestingly the L1 G users' production of *sein* passives is not 100% perfect as shown in Figure 6.3. It is interesting to note that A7 in the B corpus is one of the two respondents with a less than perfect score. The suggestion is offered that her performance in these exercises indicates that her grammar is subject to variation, which may be caused by the influence of the academic environment in which she is studying, particularly in years 1 and 2 when some cross-linguistic influence of NS English may have affected her production. Alternatively, her single error in the year 4 exercise may simply be the result of an oversight, since even native speakers make mistakes, as emphasized by the single error in the production of B4.

7.3   Agent 'by' Phrases

Von phrases with passives exist in both corpora, exemplified in (17) and (18).

(17)   Viele Gebärden   werden   in der Deutschen
        Gebärdensprache   von Mundgestiken   begleitet
        Many signs   become/are becoming   in the German
        sign language   by mouth gestures   accompanied

        Many signs in German sign language are accompanied by mouth patterns
        (Student B4)

(18)   Der englische Ausdruck (…) wird   sehr häufig   von
        österreichische Lehrer (case error)   benützt (orthographic error)
        The English expression (…) becomes/is becoming very often   by
        Austrian teachers   used

The English expression is very often used by Austrian teachers
(Student A9)

In the case of (18), however, the use of a 'by' phrase (*von*) requires DAT endings on the adjective and noun following it, which are missing, and an error occurs on the PP where the modification of the vowel *u* is incorrectly inserted. It may be that this respondent is prevented by having to process a number of grammatical elements simultaneously from producing an entirely correct utterance; this is the only example of a 'by' phrase in her exercise. Processing overload might also be the reason why respondent A3, who produces the least number of passives overall, produces no 'by' phrases at all.

7.4   Nativelike or Non-nativelike?

In this section, the extent to which the ultimate attainment of the A cohort in terms of passives can be considered NS-like will be discussed.

Having given an overview of the comparative success of L1 E learners of German L2 in using *werden* or *sein* in their own chosen context by means of the mini case study reports in Chapter 6, a more detailed analysis of the production of these learners offers a further dimension to the investigation which is conducted in this study. The questions which have to be addressed when defining acquisition were mentioned in Chapters 2.5.3.5 and 6.3; further questions relate to the level of nativelikeness which learners of a second language achieve and whether any particular features of a given L2 are unlearnable. Lardière (2013) discusses the notion of ultimate attainment, suggesting that it may be described as 'a stable state' (2013: 670) which is not necessarily nativelike. The question of how the stability of a language feature is assessed and to what extent this equates to the learner achieving nativelikeness should also be considered. Lardière (2013) remarks that the IL of proficient learners (who may nevertheless not have achieved the highest level of proficiency, such as the L1 E learners in this study) provides suitable language samples for analysis because the learners may be considered as having reached a steady state in their language development, which does not require 100% perfection. It has been indicated in Chapter 6.4 that errors may be detected in the production of NSs as well as NNSs, which allows the claim that error-free language is not required in order for an item to be considered nativelike. As far as learnability is concerned, Lardière's preferred view is the one

which Birdsong presents in his consideration of nativelike performance by L2 learners, that is, that 'no feature of an L2 is unlearnable' (Birdsong 2006: 182). This would appear to be the case illustrated in Chapter 6.3.5 and 6.3.8 where the CCs of respondents A5 and A8 are tabulated. 100% accuracy of choice is recorded for both of these two for both types of passive, although the incidence of *sein* passives is very small (6 and 3 tokens respectively) and therefore cannot be considered as a representative sample.

In a consideration of the acquisition of morphosyntax by adult L2 learners, Ionin (2013: 505) indicates the number of studies in the generative tradition which have considered whether learners are capable of achieving a 'targetlike syntactic representation, especially in those domains where the learners' native language and their target language differ'. For two reasons, the study described here appears to sit within this spectrum: firstly, 'targetlike' here relates to a language type which, for assessment purposes in a pedagogical environment, is as close to the one specified in grammar textbooks as possible, as indicated in Chapter 2.6.2. Secondly, in Chapter 3, the difference between expressing the periphrastic passive in German was contrasted with the 'be' passive in English where only one auxiliary is required to perform the passive function and this difference must be mastered by L1 E learners of German. The majority of the studies which Ionin mentions give the cause for deficiencies in the acquisition of morphosyntax as the complexity of representing syntax, inflectional morphology and/or phonology, but additionally Ionin refers to a proposal by Lardière offering a different perspective, the Feature Reassembly Hypothesis (Lardière 2009). This considers whether particular items are acquirable from both a syntactic and a semantic point of view and supports the claim, already mentioned, that there are no items which cannot be acquired by L2 learners. From the evidence presented in Section 7.2 and in the previous qualitative analysis in Chapter 6, it can be seen that while the L1 E learners of German in this study produce nativelike passive morphology with few errors, their selection of the pragmatically suitable passive auxiliary is still defective in some cases. In the next section, alternatives to the periphrastic passive which appear in the A and B corpora are considered, in order to present a fuller picture of the use of the passive by both native speakers and L1 E learners of German in the formal environment in which this study is set.

7.5 Alternatives to the Passive

It is known that use of some previously learned structures declines as other, alternative constructions with a similar meaning are acquired (e.g. Richards 1987 for L1; Ellis 2013 for L2). It is possible that this applies to the German passives produced by learners as time progresses; the use of alternatives, explained in Chapter 4.8.2.1, will now be considered.

Certainly the calculation of passives-per-500-words illustrated in Table 5.18 indicates a decline in token frequency of the auxiliaries for six of the eight respondents over four years in the A corpus. A1 and A3 are the exceptions. In Chapter 6.6, the production of two alternatives to the passive, *sein +zu* + INFIN and a construction involving the use of the verb *sich lassen* + INFIN was examined. It was noted that there are no tokens of either construction in the Year 4 exercise in the A corpus, in comparison to 88 and 25 tokens respectively in the B corpus.

7.5.1 *Man* + a Verb in the Active Voice

The pronoun *man* (= one; someone; we) appears 158 times in the A corpus, 34 times in sentence initial position. An SVO word order similar to the usual word order in English would seem to be being observed in these cases.

(19) Man   kann   die steigende Zahl   im Diagramm (…)
      sehen
      One   can (3pers sing)   the increase   in the diagram (…)
      see (INFIN)

      The increase can be seen in the diagram/We can see the increase in the diagram
      (Student A1)

There is a preference for *man* to be followed by a modal verb in the A corpus, as seen in constructions such as a conditional (20).

(20)   Man   könnte   auch   sagen

        One   could (3pers sing conditional)   also   say (INFIN)

        It could also be said

In contrast, there are only 42 instances of *man* in the entire B corpus, 16 produced by B3, two by B1, eight by B3 and 16 by A7. There is no evidence at all of the use of *man* by B4, albeit in a shorter exercise than the other B respondents. From this we surmise that the L1 E learners opt for an active construction much more frequently than the NSs of German in order to avoid the complex demands of producing a passive.

In the B corpus, there is a noticeable number of instances of the pronoun *man* being used with an inverted verb initial form where a clause introduced by *wenn* (= if) would have been acceptable, for example (21) and (22).

(21)   Berücksichtigt   man   die oben angeführte   Kontakthäufigkeit

        Considered (3pers sing)   one   the above mentioned   frequency of contact

        If the above-mentioned frequency of contact is taken into consideration

        (Also possible: Wenn man die oben angeführte Kontakthäufigkeit

        berücksichtigt = If we consider the frequency of contact mentioned above)

        (Student B3)

        Variation in sentence structure is also seen with the inversion of *man* and the verb in the B corpus (22).

(22)   Darunter   versteht   man   starke   lineare
        Abhängigkeiten

        Among these   understands (3pers sing) one   strong   linear
        dependencies

        Strong linear dependencies are seen here

        (Student B1)

In fact, *man* appears rarely in sentence or clause initial position in the B corpus, a total of three times, two of these being the option for A7. Once again, it may be observed that her use of an item is different from that of the other respondents in the B corpus. It is possible that she may be exhibiting some influence from English L1 and following a pattern used by L1 E learners, but such a suggestion would necessarily have to be confirmed by questioning the respondent in person before any claim regarding attrition could be made.

7.5.2   *Sich lassen* + **INFIN**

Use of a second alternative to the passive, a 3pers form of the verb *sich lassen* (to allow something to be done) + INFIN has already been discussed in Chapter 6.5 where a longitudinal analysis of the entire A corpus was carried out, rather than an examination of the Year 4 exercise only which is the subject of this chapter. The point was made there that this construction does not appear in the Year 4 exercises in the A corpus, but conversely appears albeit to a limited extent in the B corpus. It is useful to repeat at this juncture that A corpus learners seem to use expressions with *sich lassen* in a formulaic way and to stop using them as time progresses.

7.5.3   *Sein* + *zu* + **INFIN**

As previously discussed in Chapter 6.6, this construction, which is an alternative to a *werden* passive, occurs far more frequently in the B corpus than in the A corpus where its use is confined almost entirely to the year 1 exercise. Thus in the context of a comparison between the A corpus year 4 exercise and the production of the B cohort, there are no examples to cite from the former.[4] (23) and (24) are to be found in the B corpus. These respondents appear to be using a greater variety of passive equivalents than those with L1 E.

(23)   Während dieses Gütemaß   nur schlecht   zu veranschaulichen   ist
       While this volume   only with difficulty   to visualise   is

       While such a large amount is difficult to visualise

---

[4] Examples from the year 1 exercise in the A corpus are given in Chapter 6.6.

(Student B1)

(24)   […] ist in diesem Rahmen   nicht   zu klären
         […] is in this context   not   to clarify

         […] is not to be clarified in this context
         (Student B4)

Although, as indicated in the introductory paragraph to Section 7.5, use of some structures declines as learner language develops, the production of other structures increases as the number of occasions on which they are encountered increases (Sharwood Smith, Truscott & Hawkins 2013). It is possible that NSs of German have had more exposure in their native-speaker environment to constructions with *sich lassen* and *sein + zu +* INFIN and that lack of exposure accounts for the lack of production of these constructions by the A cohort.

7.6   Concluding Remarks on Chapter 7

In this chapter, the production of the passive by the respondents in the A cohort has been examined in relation to patterns observed in the B corpus.

Having provided evidence to indicate that learners of German with L1 English acquire both the function and the grammatical form of the *werden* passive before the *sein* passive, it seems straightforward to conclude that instruction provides the basis for production, since the *werden* passive is taught as 'the passive' to L1 E learners in formal learning environments as indicated in Chapter 3.6. However, examples of the *sein* passive would have been given in contextually relevant situations in a formal learning environment to provide additional support for instruction, in the expectation that acquisition would be promoted by a combination of this with the reading of texts and participation in spoken activities. It is therefore more difficult to advance a reason for the non-nativelike patterns in the production of the *sein* passive which are considered in section 7.2. It is suggested that these are primarily the result of negative transfer from the learners' L1. It was shown in Chapter 6.3 that, according to the criteria adopted in this study, in the A corpus one respondent, A5, can be said to have acquired the *sein* passive, and it was pointed out that that particular student could be

regarded as a cautious learner because use of the construction, although just achieving the required benchmark, is relatively limited. This gives support to the suggestion in Section 7.4 that there are no items that cannot be acquired by learners. This does not mean that every learner will ultimately produce nativelike language but that some will while others may exhibit permanent fossilization.

Some L1 E learners perceive the German passive more generally to be problematic and therefore deliberately avoid it because they are aware that the statal passive exists and is different from what they have been taught is the passive (i.e. *werden*) but are not confident enough of their own ability to select the appropriate grammatical form for the context they have chosen. A preliminary interpretation of the questionnaire (Appendix A) reveals that respondents A3, A6 and A9 admitted in their positive response to question 3 that they tried to avoid using passives in German.

As already stated in Section 7.2.2, the introduction of a think-aloud protocol or a stimulated recall exercise, for example, would help to account more accurately for non-native-like usage of the passive.[5] Future research such as this will be directed towards further analysis of the supporting data collected in the questionnaire and the GJE (Appendices A and B). At this stage of the research, the results of this survey indicate both inter- and intra- learner variety in the production of German passive constructions and similar analyses of a larger corpus are necessary before any conclusions can be applied to the L1 E learner population in general.

In Chapter 8, the conclusions which can be drawn from the quantitative and qualitative analyses will be used in an estimation of the relative success of this study in terms of the goals which it set out to achieve. The relevance of the findings to second language pedagogy will be considered.

---

[5] The two methodologies are described in detail by Gass (2013: 49-50).

CHAPTER 8 – REVIEW AND CONCLUDING DISCUSSION

8.1   Introduction

The concluding chapter of the thesis begins by reviewing the structure of the study and the implementation of the methodology. This is followed by a restatement of the research questions in order to assess the extent to which they have contributed to the achievement of the research aims. The contribution made by this study is assessed.

8.2   Review of the Research Structure and Implementation of the Methodology

Following the introductory Chapter 1, Chapter 2 introduced definitions of the passive found in the academic literature and a survey of some relevant studies in SLA with particular reference to the relationship of teaching to that discipline. A survey of the development of corpus studies and of learner corpora followed, relating the data collection approach to the academic disciplines. In Chapter 3, a more detailed explanation of the German passive was related to pedagogical practice.

Chapter 4 identified the precise methodology which was used to examine the data collected in the LINCS corpus. Both a quantitative and a qualitative analysis of a corpus of learner German were carried out, as described in Chapters 5 and 6, together with a similar approach examining a corpus of NS German and a corpus comprising a limited amount of textual material in German which had been used in teaching activities. This additional dimension was included in the evaluation of the results of the analyses, looking for support for an initial hypothesis related to the sequence of acquisition of the various ways of expressing the passive in L2 German in a formal learning environment. A more detailed qualitative analysis of the NNS speaker corpus was reported in Chapter 7 in order to investigate more precisely how L1 E learners use the German passive and to compare this with NS production in a similar exercise.

Learner corpus research as a genre has not frequently been applied to the acquisition of second and subsequent languages and Granger (2012: 9) criticizes the fact that where it has been applied '[t]here is rarely any information on the teaching methods, the course material or the first language (L1) or L2 status of the teachers'. In order to reduce still further the variables inherent in pedagogical research, these factors have all been

mentioned in this study and thus offer an additional layer of information with which to support a diverse methodological approach.[1] As Ortega (2003) has pointed out, more consistency is needed in order to draw reliable comparisons between the conclusions reached in various studies.

Norris & Ortega (2003) pointed to the possibility of raters' scores being influenced by, among other things, subjective bias in identifying and subsequently encoding the phenomenon under investigation. Having decided to assess the learners' performance on the basis of correct/incorrect in terms of the use of the auxiliary verb *werden* or *sein*, there is evidence to justify the claim that the original rater's decisions were not influenced by biased expectation. The possibility of error in categorising 'passive' or 'not passive' in the concordancing stage should not be ignored; fatigue, frustration, boredom and what Norris & Ortega neatly term 'drift' (Norris and Ortega 2003: 743) were all experienced during that phase of the study, but this was counteracted by varying the length of time spent on that particular activity. In addition, most of the PASS[2] files were reviewed after the preliminary concordance for the purposes of the longitudinal analysis. A few errors were identified but they were limited in number and were corrected at that stage.

According to the early hypothesis, based on a preliminary assessment of the written production of the A cohort, *sein* passives are acquired less easily by learners with L1 E than *werden* passives. Whilst the accuracy of the judgment of a bilingual[3] teacher of languages with German L2 might be called into question, an NS of English, equipped with the linguistic knowledge of forming the passive in English, should be well placed to identify 'passive intent' in an L1 E learner of L2 German, since the instructor will have been faced with the same grammatical obstacles in learning how to use the *werden* and the *sein* passives him or herself. Nevertheless, a measure was put in place to ensure that the judgement on which the hypothesis rests is correct. The views of one rater with L1 G (Appendix D) and one with L1 E (Appendix E) provided support by assessing a number of instances of the passive produced by the B cohort and the A cohort respectively and judging the semantic correctness of each one.

---

[1] Teaching methods are described in Chapter 1.2, course material in 3.6 and the L1 and L2 status of the teachers in 4.6.1.1.
[2] See Chapter 4.8.2.1.
[3] 'Bilingual' as defined in Chapter 1 footnote 4.

The learner corpus has proved useful in enabling the *count* and *sort* procedures to be carried out. The larger the corpus, the more likely it is that results can be reliably extrapolated to a wider population of learners. The size of the data sample in this study is relatively small, but it is the case that similar studies have been usefully carried out in pursuit of a contribution to SLA theory (e.g. Hikima 2010 with only 10 subjects). Since it is true that statistical results based on small numbers can give an unintentionally biased perspective, an effort must be made to ensure that analyses take careful account of as many variables as possible and anomalies are pointed out. Student B4's score of 75% CC for *sein* passives after administering the formula adopted for this study is based on her making one error in only 4 tokens, whereas A7 makes one error in 28 tokens, giving her a score of 96.4% CC.[4] The point has been made, however, that even native speakers make mistakes and a perfect score is therefore neither required nor expected in order to indicate acquisition. Determining the point of acquisition will depend on the researcher's stance, as illustrated in Chapter 2.5.3.5.

The combination of measurement tools was implemented to ensure as robust an outcome to the analysis as possible.

8.3   Responses to the Research Questions

While the general aims of the study were to investigate patterns in the usage of the German passive and subsequently to suggest possible reasons for these patterns, more precise research questions were formulated to provide the framework for the quantitative analysis in this study. The first three were expressed as follows:

1. Which forms of the German passive are produced in written interlanguage by adult learners with L1 E?
2. Do these forms vary over time?
3. Which forms of the passive are evident in the written production of adult NSs of German?

The learner corpus provided a valuable data source of examples of passive usage and indeed of lack of passive usage, enabling answers to be provided to these three questions.

---

[4] See Chapter 6.2.

On the basis of the results of the *count* and *sort* procedures, a *comparison* between the evidence from the study of the patterns in the learner language and those of the NSs of German was possible and is reported in Chapters 6 and 7, in response to research question 4.

4. What similarities or differences can be observed between the use of the passive by NSs and NNSs in the respective subcorpora?

It was pointed out in Chapter 6.4 and 6.7 that NS users of German not only produce a wider variety of passives than NNSs but also that they are more accurate in their production of the two types of periphrastic passive.

Across the A cohort, only A5 and A8 match the form and function of their target *sein* passives in year 4, whereas every one of the other respondents makes a correct match at some other juncture in their studies but fails to do so in year 4. This type of variation, where NS-like forms which have previously been recorded seem sometimes to disappear, only to re-emerge later, accords for example with the prediction of the MOGUL framework,[5] that both types will coexist until the target form becomes dominant. It seems that even though the function of basic syntax may have been acquired (the learners know that a passive is required), the learner appears to be taking a step backwards by opting for a statal passive form where an action is being described. In this context, the learners' usage of the passive often diverges from the NS norms while at times converging with those norms, and thus demonstrates incomplete mastery of the construction, to use Papp's (2000: 177) terminology which was mentioned in Chapter 2.5.3.7.

Where *werden* passives are concerned, all of the members of the A cohort except A3 produce contextually correct *werden* passives in year 4, with three of them producing accurate matches at all the levels for which data were recorded. There are no data available for A4 in year 1, however, and A6 produced no passives in year 2 in the examination answers. Even if the relatively low number of instances in some cases does not indicate acquisition as it has been defined in Chapter 2.5.3.5, it seems reasonable to remark that production of the *werden* passive by the L1 E learners of L2

---

[5] See Chapter 2.5.3.1.

German is more stable after 4 years of university-level study than is their production of *sein* passives.

Other perspectives may also account for the lack of accuracy in the production of the two passive types in the L1 E corpus. Sorace (2005), for example, makes a distinction between soft and hard structures, of which the former involve the interaction of syntax and other domains such as semantics or pragmatics whereas the acquisition of the latter involves the knowledge of syntax only. Soft structures, therefore, might be expected to be problematic for learners such as those with L1 E who must acquire the difference between the form and the meaning of the two German periphrastic passives. Sorace & Filiaci (2006), in introducing the notion of the IH mentioned in Chapter 2.5.3.7, go further by predicting that at the interface between syntax and other modules of the grammar, non-native items may continue to appear even after the basic (hard) syntax is acquired. In the case of the L2 G learners in the study reported here, the fact that *sein* passives which are morphologically correct are still used variably in the incorrect semantic context in year 4 supports the predictions of the IH. This contention is further supported by Rankin (2011: 139) who indicates that '[in] production (…), it is to be expected that near-native speakers of an L2 will optionally produce non-target forms in interface areas alongside target L2 forms.'

8.4   Achievement of the Research Aims

Two research aims were identified at the beginning of the study, the first of which is repeated below.

> To contribute to the theoretical discussion concerning the relationship between SLA and second language pedagogy.

8.4.1   Application of the Findings to L2 Pedagogy

To what extent can it be said following this study that instruction affects the acquisition of the passive in a formal learning environment? Although studies have shown that children learning German as L1 acquire the *sein* passive in advance of the *werden* passive (e.g. Abbot-Smith 2003), the situation in a formal L2 learning environment is different. As indicated in Chapter 3.5, the relatively rare occurrence of the *sein* passive

in contemporary German texts suggests that it will not appear frequently in the input. The language teacher's task is to equip learners with as complete a range of FL structures as possible and supplement the examples in grammar textbooks such as Durrell (1996) which offer explanations that learners may find difficult to relate to everyday usage.[6] To make the construction more salient, additional examples will be supplied with the intention of scaffolding acquisition with more frequent and contextually relevant evidence and it is therefore surprising that the *sein* passive in this data sample is only seen to be 'acquired' by one of the L1 E respondents, as indicated in Chapter 7.6.

In tertiary education, where knowledge, if not control, of the passive is expected, the terminology used to describe the *sein* passive is variable. The term *statal* was selected in this study as an English alternative to the German term but it might be that *adjectival* would be more useful. Conversely, given that Duden points out that *sein* passives are not to be confused with other constructions formed with that verb, among which are adjectival expressions (Duden 1998: 185), explanations involving only the term *adjectival* would not seem to offer a complete answer. In the case of subjectless or Duden Type C passives, learners might produce grammatical utterances relatively easily if a construction is taught as a formula (Sharwood Smith 2013), in part because of the reduction in processing effort which reproduction of a prefabricated expression affords (Wray 2002: 18). The study described here has not suggested novel ways of teaching the *sein* passive; an experimental study comparing groups who are offered different explanations of this construction would constitute an extension to the outcomes. A future experimental study considering the written production of learners with L1s other than English would also be an obvious successor to this project, since it may well be the case that the sequence of the acquisition of the *werden* passive and the *sein* passive in L2 is different in the case of languages other than English.

8.4.2   Instructional Input

Various types of instructional input have been considered in the study reported here (e.g. Chapter 2.5.3.1). The *werden* passive is taught at secondary school level as 'the passive' to learners with L1 E. Indeed, the *English grammar for students of German*, mentioned in Chapter 3.2, produced for a readership at upper-secondary and early

---

[6] See Chapter 3.7.

university level, avoids mentioning the *sein* passive. For Zorach & Melin (1994), at least, the construction is considered either unnecessary or too complex for their target audience. Usage of the *sein* passive by learners with L1 E may be governed at least initially by the lack of prominence of the construction in the pedagogical input.

With the exception of the six-month spell abroad, from which no evidence of input is available, all of the students will have had more or less the same recourse to grammar text books, written and spoken FL input and formal instruction in a classroom environment. They will have received explicit instruction on the formal construction of the two types of syntactic passive during their first year of study, but in second year they will have received little explicit grammar instruction and will have had to seek out references for grammar reinforcement themselves. How and to what extent they used that varied input will depend on the individual. Explicit feedback on their production throughout their university course will have been provided in both oral and written form but it is not possible to quantify the benefits which can be attributed to it. The aim here has been to observe the respondents' written production of the passive up to a certain point in year 4 and identify the patterns in its use over preceding years.

It would appear from the analysis of the LINCS corpus that for some L1 E learners of German the production of the *sein* passive will remain unstable in their written interlanguage. A review by Lardière of nativelike and non-nativelike attainment concludes that 'most endstate L2 learners (…) do not [achieve nativelikeness in all respects]' (Lardière 2013: 691). There is no evidence in the corpus study to suggest that there will be a positive and long-lasting effect on the written German (with which this study is concerned) of the learners in the L1 E cohort brought about by the type of instruction which the learners receive. This is perhaps one of the 'respects' to which Lardière refers. While the influence of L1 English may subconsciously be a cause of continuing NNS-likeness for some L2 G learners of the *sein* passive, as suggested in Chapter 2.5.3.7, 2.5.3.9 and in 2.6.1, it is also possible that the learners simply lack sufficient practical writing experience, despite communicative teaching methods and a period of study abroad, to guarantee high levels of attainment in using the German passive in a nativelike way both in its form and in its functional contexts. The evidence from the type of corpus study research reported in this thesis gives an indication to teachers of the extent to which their expectations of learners' achievements are likely to be met, and thus may ultimately have an influence on methods and levels of assessment.

Since the second of the research aims of the study reported here, as stated in Chapter 1.7, is to make the data available to a wider research population by offering it for inclusion in an existing, more extensive database of learner interlanguage, it is possible that future research using similar methodology will enable more generally applicable conclusions to be reached concerning the acquisition of the German passive. On the other hand, even a larger data sample may not provide sufficient evidence to enable definitive claims to be made because the construction, particularly the *sein* passive, does not occur frequently in the input and even the added support for acquisition supplied by instruction may not result in acquisition in all cases.

To sum up, considering the frequency of occurrence of both constructions, the *werden* passive is likely to remain dominant as 'the passive' to learners of German with L1 E. Ultimately it is probable that students will acquire the *sein* passive only when and if their mental grammars are sufficiently well-developed to do so.

8.4.3   Sharing Data with a Wider Research Population

With reference to research aim 2, identified in Chapter 1.7, to construct an embryonic learner corpus and offer data to a wider research population, positive progress is reported. In the short term, this relatively small corpus will be offered to a wider research public as an addition to an existing collection. The LINCS corpus is already listed on the UCL website and recently another researcher has requested permission to access it. For the reasons given in Chapter 4, the current corpus remains unannotated, but annotation is an option for future research with a different orientation, for example morphological errors. The data from this study can be offered to an existing collection of learner data for PoS tagging[7] or for error tagging.[8] Uploading is a relatively simple process and is one of the clear advantages of electronic databases as was indicated in Chapter 2.6.3. Brian MacWhinney and Paul Schmidt have already indicated their willingness to accept the LINCS corpus, but the terms of access have not yet been investigated. In addition, a new collection of Instruments for Research into Second Languages (IRIS) has been established at the University of York which allows materials or measurement instruments to be uploaded so long as they have been used to collect data in published, peer-reviewed work or in an approved PhD thesis. The responses to

---

[7] See Chapter 4.4.2.
[8] See Chapter 4.4.1.

the questionnaire[9] and the GJE[10] which have already been completed by the A cohort will be cross-referenced with the corpus data of the learners' production of the passive, and the two measurement tools could then be offered to this collection.

8.5   Conclusion

With respect to the achievement of the general research aims, this study has investigated patterns in the written production of the German passive by adult learners with L1 English in a formal learning environment and uncovered some interesting facts. It did not set out to test a particular theoretical stance but to observe patterns in the acquisition of the construction through the analysis of a learner corpus and to relate those patterns to selected second language acquisition theories. After investigating the acquisition of the construction in L1, it was revealed that the acquisition order of the *sein* and *werden* passives differs in the present context in L2 acquisition. A further comparison with the written production of adult NS users of German indicated that the NNSs of German in the corpus made different choices when using passives in comparison to NSs, the latter group being capable of expressing the passive in alternative ways.

As a result of this cross-disciplinary study, some reasons which might account for the differences between NNS and NS performance have been suggested. Others, such as learner strategies in the use of the passive, are worthy of future exploration. Examining data retained in a learner corpus can facilitate the analysis of learner language in a response to the challenges of further research.

---

[9] Appendix A.
[10] Appendix B.

APPENDIX A

QUESTIONNAIRE FOR FOURTH YEAR STUDENTS

Please do not use any reference works in order to answer the questions – just complete them at one attempt, with minimal reflection. Once you have completed a question, please do not return to it and change the answer – I am interested in your *first* reaction. There is no need to read the whole thing through beforehand; indeed it is better if you do not. A page break has been inserted after question 10 to prevent you from looking at the next question too soon!

I have asked you to give me your name so that I can tally your questionnaire with your written work – but you will **not** be identified by name in the research report.

For the purposes of the questionnaire, these terms and abbreviations are used:

auxiliary = *sein* or *werden*

pp = past participle

subject = subject of verb i.e. Nominative case

object = object of verb i.e. Accusative case

active = verbs conveying an active, as opposed to a passive, meaning

Please indicate your preferred answers by checking the boxes at the end of the answer(s) you have chosen. Don't be put off by the first question – it is the only theoretical one.

Notes

i. Clicking the box twice removes the x

ii. For typed text responses, click in the larger grey box, which will turn black and expand as you type

Your name:

Please fill in the time now (hours and minutes e.g. 10.43):

1. What does *passive* mean to you? Choose one.
   Is it:
   a) a construction which focuses attention on the person or thing carrying out the action of the verb? ☐
   b) a construction which focuses attention on the person or thing affected by the action of the verb? ☐
   c) a construction which focuses attention on the action of the verb itself? ☐
   d) something else ?(please say what)

2. When you use a passive construction in German, which of the following do you personally give most thought to? Choose one.
   a) the form of the pp ☐
   b) the choice of auxiliary ☐
   c) the change of object/subject necessitated by the passive ☐
   d) the form of the auxiliary ☐

3. Would you say that you try to avoid using the passive in German?
   a) Yes   ☐
   b) No   ☐

4. If you answered *yes* to the previous question, can you say why?

5. What would you be most likely to choose if you did decide to use something other than a passive construction?
   a) a construction with *man*   ☐
   b) an active construction using the same verb   ☐
   c) an active construction using a different verb ☐

6. If you chose option (c) in the previous question, why would you use a different verb?

Can you give a quick example? If not, say *no*.

7. How frequently do you use *sich lassen + infinitive* with passive meaning? (e.g. Das lässt sich auf folgender Weise erklären… {*That can be explained in the following way…*})
a) whenever I think it fits ☐
b) sometimes ☐
c) never, because I'm not sure how it works ☐
d) never, because as far as I know I've never come across it ☐

8. Can you give an example in German of what is known grammatically as a "subjectless passive"?
   a) No ☐
   b) Yes; here's an example:

9. Which of these best describes the way you aim to use the correct case with verbs? (Choose more than one if you wish, but if you do, please give some idea of the order in which you would rank them by adding 1, 2 etc. in the second box.)
   a) I check most of the verbs I use, in a dictionary ☐
   b) I sometimes check the case required by verbs ☐
   c) I pay particular attention to cases with verbs when I'm using the passive ☐

   d) I check only rarely because I'm pretty good at this ☐
   e) I check only rarely because I'm a native speaker of German ☐
   f) I'm a native speaker of German so I don't need to check ☐

10. What is the difference in **function** of the *sein* passive and the *werden* passive? Check the box if you don't think you can give an explanation. ☐

If you *can* give an explanation, please write as full an answer as possible before you look at question 11.

sein:

werden:

Questionnaire continues on the next page………

11. If you did not give a written answer to the previous question, here are some examples of possible descriptions of the functions of *sein* and *werden* passives. Check the one which you now think is the best. Even if you did answer question 10, feel free to answer this one, too.

   a) *sein* + pp represents an action in process and *werden* + pp represents a state
      ☐

   b) *sein* + pp represents a state and *werden* + pp represents an action in process
      ☐

   c) *ist* means *is* and *wird* means *becomes* or *gets* ☐

   d) part of *sein* is usually correct for a passive because English passive is formed with the verb *to be* ☐

   e) the pp in a *sein* passive is an adjective; in a *werden* passive the pp is a verb
      ☐

   Further comment if you wish:

12. From your experience, which do you think is more common in written German texts?

   a) *sein* passive ☐

   b) *werden* passive ☐

   c) both equal ☐

13. How would you express *There is a strike at the factory* in German?
    Check the box if you can't answer ☐

14. When you were writing your dissertation, did you at any time pay particular attention to how you formed passives?

   a) Yes ☐

   b) No ☐

15. If you answered *yes* to question 14, why was this?

16. Please rank these in order of importance, indicating what has helped you most **since you commenced your university course** to improve your production of the passive. Type a number (1-5) in the box at the end of each option.

a) spoken language in everyday use, particularly during your year abroad

b) grammar teaching

c) individual grammar practice specifically of the passive

d) reading German

e) writing German

If you would like to make any other comments, feel free to use this box:

Please fill in the time now (hours and minutes e.g. 11.02):

Now save the document and return it to me as an email attachment. Please save it as:

Yournamequestionnaire.dot

e.g.EThodayquestionnaire.dot

Once again, I am extremely grateful to you for giving up your time to cooperate on this project.

APPENDIX B

Grammaticality judgement exercise

Please check the box at the end of each sentence if you think it is *grammatically correct*. If you think it is incorrect, no mark is necessary. Finally, there is a grey type-in box in which to reflect on your thoughts. As before, I need your name simply in order to be able to link your various responses together. Don't think for too long; try to react intuitively. (If you decide to change your mind, clicking on the check box a second time removes the x.)

The time now:

Your name:

1. In Schottland darf man nicht mehr in öffentlichen Gebäuden rauchen. ☐

2. Mißbräuchliche Benutzung wird bestraft! ☐

3. An meiner Party werden wir alle viel getrunken und gegessen. ☐

4. Das Fahrrad ist repariert. ☐

5. In der Kirche wird es gesungen. ☐

6. Mitte Januar wird das Wetter immer kälter geworden. ☐

7. Es wird nach dem Fussballspiel gefeiert. ☐

8. Während der Komödie werden die Zuschauer oft gelacht. ☐

9. Es wird in der Fabrik gestreikt. ☐

10. Ein neues Studentenwohnheim wird es hier gebaut. ☐

11. Morgen wird es kühler sein. ☐

12. Er ist mit Staub bedeckt. ☐

13. Der Artikel wird in einer Fachzeitschrift veröffentlicht. ☐

14. Auf der Strasse wird getanzt. ☐

15. Ein neues Theaterstück hat der Münchener Regisseur aufgeführt. ☐

Please tell me what you were thinking while you were reading these sentences -just write spontaneously:

The time now:

Now please save the document as yournamegrammexo.dot and send it back to me.

Und nochmals recht herzlichen Dank!

APPENDIX C

PERMISSION TO USE DATA

I do / do not (*please delete as applicable*) give permission for staff and students in the Department of Languages and Intercultural Studies (LINCS) at Heriot-Watt University to use examples of my spoken and written student performance for research purposes, on condition that all the data are anonymised and that no individual can be identified in any published results, either internally within the university or as part of an external project.

Signature……………………………………………………

Name (Please print in block capitals)…………………………………………………..

Year of Entry: 20…….

Course of study (*please delete as applicable*):

Undergraduate: LINT/ALT/IBML/FALS/GALS/SALS/MEL/CEL/Exchange

Postgraduate: MSc / PhD

Personal data recorded for teaching, assessment and research will be held securely by your School for a limited time in accordance with your rights under the Data Protection Act, 1998 and the University's records retention policy and destroyed confidentially when no longer required. Your School may retain examples of assessed work for a longer period to meet the review requirements for professional bodies. The University keeps a limited, permanent, research archive containing records of lasting research value. For further information please contact your course administrator or the University's Data Protection Officer.

Inter-rater-assessment: passives (L1 G)

Do you, as a native speaker of German, consider the following instances of the verbs *sein* and *werden* to be syntactically correct as examples of an intention to express passive (*Zustandspassiv* or *Vorgangspassiv*):

	Please ignore
1. Je größer der Abstand der Marktanteile eines Unternehmens zu seinen Mitbewerbern schließlich ist, desto wahrscheinlicher ist, dass ihm ein unabhängiger Verhaltensspielraum zukommt und damit eine Marktbeherrschung gegeben **ist** yes ☐ no ☐ not sure ☐ comment	B2 (28)
2. wovon nach dem Zusammenschluss nicht mehr auszugehen **ist** yes ☐ no ☐ not sure ☐ comment	B2 (32)
3. Dafür **ist** eine Abwägung zwischen den durch den Zusammenschluss zu erwartenden Verschlechterungen der Wettbewerbsbedingungen und den durch den Zusammenschluss zu erwartenden Verbesserungen der Wettbewerbsbedingungen vorzunehmen yes ☐ no ☐ not sure ☐ comment	B2 (35)
4. dass das Wettbewerbsgeschehen erstarrt **ist** yes ☐ no ☐ not sure ☐ comment (please add whether you consider this would also be acceptable as a *werden* passive (*erstarrt wird*)	B2 (46)
5. Wenn die am Zusammenschluss beteiligten Unternehmen auf demselben Markt tätig und mithin Wettbewerber sind, **ist** davon auszugehen, dass es zu einer Addition der Marktanteile kommt yes ☐ no ☐ not sure ☐ comment	B2 (51)
6. Folgerichtig stellt der Erwerb der Anteilsquote, bei der von einem Unternehmenskauf als „share deal" auszugehen **ist**, neben einem Kontrollerwerb bereits einen Anteilserwerb dar. yes ☐ no ☐ not sure ☐ comment	B2 (57)
7. Vorausgesetzt, die Schwelle des Art. 1 II FKVO **ist** nicht erreicht, hat ein Zusammenschluss ebenso gemäß Art. 1 III FKVO gemeinschaftsweite Bedeutung yes ☐ no ☐ not sure ☐ comment	B2 (104)
8. beispielsweise die Enttäuschung darüber, dass der Partner 'real' ganz anders agiert und eingestellt **ist**, als es beim Online-Kennenlernen den Anschein gemacht hat yes ☐ no ☐ not sure ☐ comment	B3 (4)
9. Auffällig ist hier allerdings, dass 91,4 Prozent der Nicht-Dater, die noch in einer Beziehung leben, die Kontakthäufigkeit mit mindestens einmal täglich angeben, aber nur 66,7 Prozent der Nicht-Dater, bei denen die entsprechende	B3 (12)

Beziehung schon beendet **ist**. yes ☐ no ☐ not sure ☐ comment	
10. Wie genau aber die Umstände des Online-Kennenlernens aussehen **ist** – besonders im deutschsprachigen Raum – bisher allerdings wenig erforscht. yes ☐ no ☐ not sure ☐ comment	B3 (27)
11. Ob es einen Unterschied im Kennenlernprozess und der Beziehungsentwicklung gibt, **ist** dabei bislang nicht geklärt. yes ☐ no ☐ not sure ☐ comment	B3 (60)
12. Durch eine leicht modifizierte Mimik, bei der auch die Wangenpartie etwas angezogen **ist**, wird aus der zuvor neutralen Frage, ein leicht anzweifelndes Fragen yes ☐ no ☐ not sure ☐ comment (please add whether you consider this would also be acceptable as a *werden* passive (*angezogen wird*)	B3 (7)
13. Das Argument, dass sie der Stadt Graz Einnahmen bringt, **ist** hier nicht wirklich angebracht. yes ☐ no ☐ not sure ☐ comment	A7 (64)
14. Allerdings **sind** Frauen in den beiden jüngeren Kohorten nicht mehr signifikant benachteiligt. yes ☐ no ☐ not sure ☐ comment	B1 (21)
15. Dargestellt **sind** die Anteile in sukzessiven Ergebnisquoten. yes ☐ no ☐ not sure ☐ comment	B1 (38)
16. hierzu **sind** Beispiele angemerkt yes ☐ no ☐ not sure ☐ comment	B4 (1)
17. Nachfolgend **sind** einige der aufgeführten Punkte anhand von Beispielen erläutert. yes ☐ no ☐ not sure ☐ comment	B4 (4)
18. Oft **wird** die Gleichheit der Chancen mit gesellschaftlicher Gerechtigkeit verbunden. yes ☐ no ☐ not sure ☐ comment (please add whether you consider this would also be acceptable as a *sein* passive (*ist … verbunden*)	B1 (18)
19. Unter dem Kaufgegenstand Unternehmen **wird** eine betriebsfähige Organisationseinheit verstanden (1), die von einem Unternehmens-bzw. Rechtsträger geführt **wird** (2). 1. yes ☐ no ☐ not sure ☐ comment 2. yes ☐ no ☐ not sure ☐ comment	B2 (26) B2 (31)
20. Mittels der Profile **wird** den Nutzern ein schneller Selektionsprozess auf der Suche nach dem idealen Partner ermöglicht. yes ☐ no ☐ not sure ☐ comment	B2 (25)

21. Stattdessen **werden** sie durch das Kennenlernen ihres späteren Partners eher in einer Lebenssituation überrascht yes ☐ no ☐ not sure ☐ comment (please add whether you consider this would also be acceptable as a *sein* passive (*sind .... überrascht*)	B3 (52)
22. Hier **werden** die Augenbrauen hochgezogen, der ganze Kopf etwas geneigt und die Augen weit geöffnet. yes ☐ no ☐ not sure☐ comment	B4 (7)
23. Dieses **wurde** vom Architekten Vito Acconci primär nicht intendiert. yes ☐ no ☐ not sure☐ comment	A7 (11)

Thank you!

Inter-rater-assessment: passives (L1 E)

Do you, as a native speaker of English, consider the following instances of the verbs *sein* and *werden* to be syntactically correct as examples of an intention to express a passive (*Zustandspassiv* or *Vorgangspassiv*):

	Please ignore
1. Seit ich meiner Forschung angefangen habe, ist es klar geworden, dass das Thema von Studiengebühren sehr viele Kontroverse in Deutschland hat, und es **ist** sehr häufig diskutiert. yes ☐ no ☐ not sure ☐ comment	A1 (30)
2. Wie man sehen kann **ist** die Mehrheit der Länder, die für Studiengebühren sind, von CDU/CSU regiert. yes ☐ no ☐ not sure ☐ comment	A1 (51)
3. Die Jahrestag der Wiedervereinigung ist am 9ten November aber es **ist** nicht so viel gefeiert. yes ☐ no ☐ not sure ☐ comment	A3 (73)
4. Trotz alles, eine Bericht in die Deutsch Welle **ist** geschrieben, dass die Lücke im Wirtschaft zwischen Ost und West ist nicht so groß. yes ☐ no ☐ not sure ☐ comment	A3 (99)
5. Die Österreich **ist** nicht normalerweise als Einwanderungsland genannt yes ☐ no ☐ not sure ☐ comment	A4 (8)
6. Erstens war diese Quota 10% der Arbeitsbevölkerung, aber nach die Eröffnung der Grenzen innerhalb der Europäische Union **ist** die Quota bis 9% abgesenkt. yes ☐ no ☐ not sure ☐ comment	A4 (12)
7. Erstens **ist** Arbeitseinwanderung haupsächlich um der Schlüssel Arbeiter und Industrien beschränkt. yes ☐ no ☐ not sure ☐ comment	A4 (15)
8. „Die Gebührenfreiheit des Studiums ist ein zentrales Instrument zur Herstellung und Sicherung von Chancengleichheit" aber mit dir Einführung von Studiengebühren **ist** diese Chancengleichheit verhindert. yes ☐ no ☐ not sure ☐ comment	A6 (1)
9. Die Einführung von allgemeinen Studiengebühren **ist** aber bis heute verhindert worden und ich kann nur die potentiellen Auswirkungen analysieren. yes ☐ no ☐ not sure ☐ comment	A6 (35)
10. Diese Möglichkeit, die Insel für private Veranstaltungen zu nutzen, **ist** noch nicht wirklich ausgeschöpft. yes ☐ no ☐ not sure ☐ comment	A7 (12)
11. Das Argument, dass sie der Stadt Graz Einnahmen bringt, **ist** hier nicht	A7

wirklich angebracht. yes ☐ no ☐ not sure ☐ comment	(64)
12. Selbst wenn dies nicht möglich ist, **ist** die Finanzierung der Insel ....gesichert. yes ☐ no ☐ not sure ☐ comment	A7 (28)
13. Die Sitzbänke im Inneren des Cafés sind aus blauem Kunstleder, auch die restliche Einrichtung **ist** im Wesentlichen ganz in blau gehalten. yes ☐ no ☐ not sure ☐ comment	A7 (73)
14. ...geformt wie eine schwimmende Muschel, die sich aus zwei verschobenen Hälften zusammensetzt **ist** yes ☐ no ☐ not sure ☐ comment	A7 (75)
15. So traten bereits während der Planungsphase Fehler und Probleme auf, die zwangsläufig die Frage nach sich ziehen, in wieweit das Projekt seiner natürlichen Umgebung angepasst **ist**. yes ☐ no ☐ not sure ☐ comment	A7 (103)
16. dass diese Wahrnehmung von der Politik zu dieser Zeit beeinflusst worden **ist** yes ☐ no ☐ not sure ☐ comment	A8 (11)
17. Diese **ist** innerhalb gewalttätige rechtsextreme Kreise eingebettet yes ☐ no ☐ not sure ☐ comment	A8 (51)
18. Die Dominanz von Englisch in den Sprachinitiaitven gibt einem den Eindruck, dass die Projekten eher auf wohlhabende Schüler, zum Beispiel, Kinder von Diplomaten usw. eingerichtet **ist** yes ☐ no ☐ not sure ☐ comment	A9 (28)
19. Es gibt neue Bahnstrecken und Bahnhöfe, und die Zentren vieler alter Städte **sind** restauriert **worden**. yes ☐ no ☐ not sure ☐ comment	A3 (1)
20. nur dass der Osten jetzt nicht mehr bemitleidet **wird** yes ☐ no ☐ not sure ☐ comment (please add whether you consider this would also be acceptable as a *sein* passive (*sind .... überrascht*)	A3 (2)
21. Es **wird** debattiert, ob ein System mit Aufnahmeprüfungen wirklich ein System für alle ist. yes ☐ no ☐ not sure☐ comment	A5 (3)
22. Zwei Monate später **wurden** die drei Straftäter verhaftet. yes ☐ no ☐ not sure☐ comment	A8 (7)

Thank you!

REFERENCES

Abbot-Smith, K. (2003) Piecemeal paths to grammatical productivity: how children become productive with basic event constructions. PhD thesis, University of Manchester.

Abbot-Smith, K. & Behrens, H. (2006) 'How known constructions influence the acquisition of other constructions: the German passive and future constructions'. *Cognitive Science*, 30, 995-1026.

Adolphs, S. (2007) *Introducing Electronic Text Analysis*. London & New York: Routledge.

Aijmer, K. (2002) 'Modality in advanced Swedish learners' written interlanguage' in S. Granger, J. Hung and S. Petch-Tyson (eds) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam & Philadelphia: John Benjamins, 55-76.

Bardovi-Harlig, K. & Bofman, T. (1989) 'Attainment of syntactic and morphological accuracy by advanced language learners'. *Studies in Second Language Acquisition*, 11, 17-34.

Barlow, M. (2005) 'Computer-based analyses of learner language' in Ellis, R. & Barkhuizen, G. (2005) *Analysing Learner Language.* Oxford: Oxford University Press, 335-358.

Bever, T. & Montalbetti, M. (2005) 'Noam's Ark'. *Science*, 298, 1565-1566.

Bialystok, E. & Sharwood Smith, M. A. (1985) 'Interlanguage is not a state of mind: An evaluation of the construct for second-language acquisition'. *Applied Linguistics*, 6, 101-117.

Biber, D. (1990) 'Methodological issues regarding corpus-based analyses of linguistic variation'. *Literary and Linguistic Computing*, 5, 257-269.

206

Birdsong, D. (2006) 'Why not fossilization' in Han, Z. & Odlin, T. (eds.), *Studies of Fossilization in Second Language Acquisition.* Clevedon: Multilingual Matters, 173-188.

Bley-Vroman, R. (1983). 'The comparative fallacy in interlanguage studies: the case of systematicity'. *Language Learning*, 33, 1-17.

Borer, H. & Wexler, K. (1987) 'The Maturation of Syntax' in Roeper, T. & Williams, E. (eds.) *Parameter Setting*. Dordrecht: Reidel, 123-172.

Bresnan, J. & Kaplan, R. M. (1982) 'Grammars as Mental Representations of Language', in Bresnan, J. (ed.) *The Mental Representation of Grammatical Relations*. Cambridge, MA: MIT Press, xvii-lii.

British Council (2013) *The International English Language Testing System.* Available at http://www.ielts.org/default.aspx. (Accessed 14.9.2013.)

Brown, R. (1968) 'The development of Wh questions in child speech'. *Journal of Verbal Learning and Verbal Behavior*, 7, 279-290.

Brown, R. (1973) *A First Language. The Early Stages.* Cambridge, MA: Harvard University Press.

Bruhn de Garavito, J. (2009). 'Eventive and stative passives: the role of transfer in the acquisition of *ser* and *estar* by German and English L1 speakers', in Collentine, J., García, M., Lafford, B. and Marín, F.M. (eds.) *Selected Proceedings of the 11th Hispanic Linguistics Symposium*. Somerville, MA: Cascadilla Proceedings Project, 27-38. Available at: http://www.lingref.com/cpp/hls/11/paper2200.pdf. (Accessed: 22.8.2011.)

Bryan, A. (1995) Der Erwerb des Passivs: sprach- und lernbarkeitstheorethische Betrachtungen zum Englischen und Deutschen. PhD thesis, Universität Tübingen.

Buscha, A. & Szita, S. (2007) Begegnungen: Integriertes Kurs- und Arbeitsbuch. Sprachniveau A2+. Leipzig: Schubert-Verlag.

Chalker, S. &. Weiner, E. (1998) *The Oxford Dictionary of English Grammar.* Oxford: Oxford University Press.

Chiarcos, C., Dipper, S., Götze, M., Leser, U., Lüdeling, A., Ritz, J. & Stede, M. (2008) 'A flexible framework for integrating annotations from different tools and tag sets'. *Traitement Automatique des Langues*, 49, 271-293.

Chomsky, N. (1965) *Aspects of the Theory of Syntax.* Cambridge, MA: MIT Press.

Clahsen, H. (1998) Child Language and Developmental Dysphasia: Linguistic Studies of the Acquisition of German. Amsterdam: John Benjamins.

Collins COBUILD Available at: http://www.collinslanguage.com/collins-elt-learners-of-english/cobuild. (Accessed: 2.12.2013.)

Cook, V. J. (1993) *Linguistics and Second Language Acquisition.* Basingstoke: Macmillan.

Cook, V. J. (1996) Second Language Learning and Language Teaching. 3rd edn. London: Arnold.

Cook, V. J. (2008) Second Language Learning and Language Teaching. 4th edn. London: Arnold.

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

Council of Europe (2014) *Education and Languages, Language Policy.* Available at: http://www.coe.int/t/dg4/linguistic/Cadre1_en.asp (Accessed 29.11.2014.)

Dekydtspotter, L., Schwartz, B. D. & Sprouse, R. A. (2006) 'The comparative fallacy in L2 processing research', in Grantham O'Brien, M., Shea, C. & Archibald, J. (eds.) *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006).* Somerville, MA: Cascadilla Proceedings Project, 33-40.

Djurkovic, M. (2004) 'Passive and impersonal in English and Serbian'. *Working Papers in English and Applied Linguistics*, 10, 49-101.

Dreyer, H. & Schmitt, R. (2001). *A Practice Grammar of German.* Ismaning: Hueber Verlag.

Duden (1998) *Grammatik der deutschen Gegenwartssprache*, Band 4, 6th edn. Mannheim: Dudenverlag, 172-186.

Duden online (2014) Available at: http://www.duden.de/ueber_duden/wie-kommt-ein-wort-in-den-duden. (Accessed 7.7.2014.)

Duff, P.A. (2012) 'How to carry out case study research' in Mackey, A. & Gass, S. (eds.), *Research Methods in Second Language Acquisition: A Practical Guide*. Oxford: Blackwell, 95-116.

Dulay, H. and Burt, M. (1974) 'Natural Sequences in child second language acquisition'. *Language Learning*, 24, 37-53.

Durrell, M. (1996) *Hammer's German Grammar and Usage*, 3rd edn. London: Hodder Headline.

Dušková, L. (1969) 'On sources of errors in foreign language learning', *International Review of Applied Linguistics*, 4, 11-36.

Eckerth, J., Schramm, K. & Tschirner, E. (2009) 'Review of recent research (2002-2008) on applied linguistics and language teaching with specific reference to L2 German (part 1)'. *Language Teaching*, 42, 41-66.

Edgar, M. (1932) 'The Lion and Albert', in Marshall, M. (ed.) *The Stanley Holloway Monologues* (1979). London: Elm Tree Books, 29-32.

Eisenbeiß, S. (1993) 'Auxiliaries and the acquisition of the passive in German' in *Proceedings of the Annual Child Language Research Forum*, Stanford, CA, 235-242.

Eisenberg, P. (2009) ‚Richtig gutes und richtig schlechtes Deutsch', in *Institut für Deutsche Sprache Jahrbuch 2008*. Berlin: Walter de Gruyter GmbH, 53-69.

El-Marzouk, G. (1998) *The canonical passive construction in Arabic and English*. PhD thesis, Trinity College Dublin.

Ellis, N. C. (2005) 'At the interface: dynamic interactions of explicit and implicit language knowledge'. *Studies in Second Language Acquisition*, 27, 305-352.

Ellis, N. C. (2008) 'Phraseology. The periphery and the heart of language' in Meunier, F. & Granger, S. (eds), *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins.

Ellis, N. C. (2012) 'Frequency-based accounts of second language acquisition' in Gass, S. M. & Mackey, A. (eds.) *The Routledge Handbook of Second Language Acquisition*. London, Routledge, 193-210.

Ellis, R. (1994) *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

Ellis, R. & Barkhuizen, G. (2005) *Analysing Learner Language*. Oxford: Oxford University Press.

Evison, J. (2010) 'What are the basics of analysing a corpus?', in O'Keefe, A. & McCarthy, M. (eds.) *The Routledge Handbook of Corpus Linguistics*. London: Routledge, 122-135.

Friedman, D. (2012) 'How to collect and analyze qualitative data', in Mackey, A. & Gass, S. (eds.) *Research Methods in Second Language Acquisition: A Practical Guide*. Oxford: Blackwell, 180-200.

Garside, R. (1987) 'The CLAWS word-tagging system', in Garside, R., Leech, G. & Sampson, G. (eds.), *The Computational Analysis of English: A Corpus-based Approach*. London: Longman, 30-41.

Gass, S.M. (2003) 'Input and interaction' in Doughty, C.J. & Long, M.H. (eds.) *The Handbook of Second Language Acquisition.* Oxford: Blackwell, 224-254.

Gass, S.M. with Behney, J. & Plonsky, L. (2013) *Second Language Acquisition: an Introductory Course*. 4th edn, New York, Routledge.

Gass, S.M. & Mackey, A. (2012) *The Routledge Handbook of Second Language Acquisition*. London, Routledge.

Gilquin, G. (2001) 'The integrated contrastive model: Spicing up your data'. *Languages in Contrast*, 3, 95-123.

Gilquin, G. (2008) 'Combining contrastive and interlanguage analysis' in Gilquin, G., Papp, S. & Díez-Bedmar, M. B. (eds.) *Linking up contrastive and learner corpus research*. Amsterdam, Rodopi, 3-33.

Gilquin, G., Papp, S. & Díez-Bedmar, M. B. (2008) (eds.) *Linking up contrastive and learner corpus research*. Amsterdam, Rodopi.

Gilquin, G. & Paquot, M. (2008) 'Too chatty: Learner academic writing and register variation'. English Text Construction 1, 41-61.
Available at: http://www.uclouvain.be/en-324999.html. (Accessed 14.8.2014.)

Goldberg, A. E. (2003) 'Constructions: a new theoretical approach to language'. *TRENDS in Cognitive Linguistics*, 7, 219-224.

Granger, S. (1983) The Be + Past Participle Construction in Spoken and Written English with Special Emphasis on the Passive. North Holland Linguistic Series 49. Amsterdam: Elsevier.

Granger, S. (1996) 'From CA to CIA and back. An integrated approach to computerized bilingual and learner corpora', in Aijmer, K., Altenberg, B. & Johansson, M. (eds.) *Lund Studies in English*, Lund: Lund University Press, 88, 37-51.

Granger, S. (1997) 'Automated retrieval of passives from native and learner corpora: precision and recall'. *Journal of English Linguistics* 25, 365-374.

Granger, S. (2002) 'A bird's-eye view of learner corpora research', in Granger, S., Hung, J. & Petch-Tyson, S. (eds.) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins, 3-33.

Granger, S. (2004) 'Computer learner corpus research: current status and future prospects', in Connor, U. & Upton, T. (eds.) *Applied Corpus Linguistics: A Multidimensional Perspective*. Amsterdam & Atlanta, GA: Rodopi, 123-145.

Granger, S. (2012) 'How to use foreign and second language learner corpora' in Mackey, A. & Gass, S. (eds.) *Research Methods in Second Language Acquisition: A Practical Guide*. Oxford: Blackwell, 7-29.

Granger-Legrand, S. (1976) 'Why the passive?', in Van Roey, J. (ed.) *English-French Contrastive Analyses*. Leuven: Acco, 23-57.

Greenbaum, S., & Quirk, R. (1990) *A Student's Grammar of the English Language*. Harlow: Longman.

Halliday, M. A. K. & Matthiessen, C. M. I. M. (2004) *An Introduction to Functional Grammar*. 3rd edn. London: Hodder Education.

Han, Z. (2012) 'Fossilization – A classic concern of SLA research', in Gass, S. M. & Mackey, A. (eds.) *The Routledge Handbook of Second Language Acquisition*. London, Routledge, 476-490.

Han, Z. (2013) 'Forty years later: updating the fossilization hypothesis'. *Language Teaching*, 46, 133-171.

Han, Z. & Odlin, T. (2006) 'Introduction', in Han, Z. & Odlin, T. (eds.), *Studies of Fossilization in Second Language Acquisition*. Clevedon: Multilingual Matters, 1-20.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002) 'The faculty of language: what is it, who has it, and how did it evolve?' *Science*, 298, 1569-1579.

Hawkins, J. (1986) The Comparative Typology of English and German: Unifying the Contrasts. London & Sydney: Croom Helm.

Hendrickson, R, M., Aitken, J., McGhee, J. & Johnson, A. (2010) 'What makes an item difficult? A syntactic, lexical, and morphological study of elicited imitation test items', in Prior, M. T., Watanabe, Y. & Lee, S-K. (eds.), *Selected Proceedings of the 2008 Second Language Research Forum.* Somerville, MA: Cascadilla Proceedings Project, 48-56.

Hikima, N. (2010) The effects of processing instruction and re-exposure on interpretation discourse level tasks: the case of Japanese passive forms. PhD thesis, University of Greenwich. Available at: http://gala.gre.ac.uk/id/eprint/6604. (Accessed 10.11.2013.)

Hinkel, E. (2004) 'Tense, aspect and the passive voice in L1 and L2 academic texts'. *Language Teaching Research*, 8, 5-29.

Hunt, K. (1965) 'Grammatical structures written at three grade levels'. NCTE Research Report No. 3. Champaign, IL: NCTE.

Ionin, T. (2013) 'Morphosyntax' in Herschensohn, J. & Young-Scholten, M. (eds.) *The Cambridge Handbook of Second Language Acquisition.* Cambridge, Cambridge University Press, 505-528.

Isurin, L. (2005) 'Cross Linguistic Transfer in Word Order: Evidence from L1 Forgetting and L2 Acquisition', in Cohen, J., McAlister, K. T., Rolstad, K. & MacSwan, J. (eds.) *Proceedings of the 4th International Symposium on Bilingualism*, Somerville, MA: Cascadilla Press, 1115-1130.
Available at: http://www.lingref.com/isb/4/086ISB4.PDF. (Accessed 3.7.2014.)

Jackendoff, R. & Pinker, S. (2005) 'The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky)', *Cognition*, 97, 211-225.

James, C. (2003) 'Contrastive linguistics'. Subject Centre for Languages, Linguistics and Area Studies Good Practice Guide.
Available at: http://www.llas.ac.uk/resources/gpg/2241. (Accessed 24.1.2011.)

Keenan, E. L. (1985) 'Passive in the world's languages', in Shopen, T. (ed.) *Language Typology and Syntactic Description* (vol. 1). Cambridge: Cambridge University Press, 243-281.

Keenan, E. L. & Dryer, M.S. (2007) 'Passive in the world's languages', in Shopen, T. (ed.) *Clause Structure*, *Language Typology and Syntactic Description*. 2nd edn. (vol. 1). Cambridge: Cambridge University Press, 325-361.

Klapper, J. & Rees, J. (2003) 'Reviewing the case for explicit grammar instruction in the university foreign language learning context'. *Language Teaching Research*, 7, 285-314.

Krashen, S. (1982) Principles and Practice in Second Language Acquisition. Oxford: Pergamon.

Lado, R. (1957) Linguistics Across Cultures: Applied Linguistics for Language Teachers. Ann Arbor, MI: University of Michigan Press.

Lado, R. (1964) *Language Teaching: A Scientific Approach*. New York, NY: McGraw-Hill.

Lardière, D. (2009) 'Some thoughts on the contrastive analysis of features in second language acquisition'. *Second Language Research* 25, 171-225.

Lardière, D. (2013) 'Nativelike and non-nativelike attainment', in Herschensohn, J. & Young-Scholten, M. (eds.) *The Cambridge Handbook of Second Language Acquisition.* Cambridge, Cambridge University Press, 670-691.

Larsen-Freeman, D. & Long, M.H. (1991) *An Introduction to Second Language Acquisition Research.* London: Longman.

Lightbown, P. M. (1998) 'The importance of timing in focus on form', in Doughty, C. & Williams, J. (eds.) *Focus on Form in Classroom Second Language Acquisition.* Cambridge: Cambridge University Press, 177-196.

Lightbown, P. M. & Spada, N. (2011) *How Languages are Learned.* Revised edn. Oxford: Oxford University Press.

Lightbown, P. M., Spada, N. & White, L. (1993) (eds.) 'The role of instruction in second language acquisition'. *Studies in Second Language Acquisition*, 15, 143-259.

Long, M. H. (1996) 'The role of the linguistic environment in second language acquisition', in Ritchie, W. & Bhatia, T. (eds.) *The Handbook of Second Language Acquisition*, San Diego, CA: Academic Press, 413-468.

Long, M. H. & Robinson, P. (1998) 'Focus on form: theory, research and practice', in Doughty, C. & Doughty, W. J. (eds.) *Focus on Form in Classroom Second Language Acquisition.* Cambridge: Cambridge University Press, 15-41.

Lozano, C. & Mendikoetxea, A. (2013), 'Learner Corpora and Second Language Acquisition: The design and collection of CEDEL2', in Ballier, N., Díaz-Negrillo, A. & Thompson, P. (eds.) *Automatic Treatment and Analysis of Learner Corpus Data.* Amsterdam: John Benjamins.

Lüdeling, A. (2006) ‚Das Zusammenspiel von qualitativen und quantitativen Methoden in der Korpuslinguistik', in Zifonun, G. & Kallmeyer, W. (Hrsg.) *Jahrbuch des Instituts für Deutsche Sprache 2006.* Berlin: de Gruyter, 28-48.

Lüdeling, A. (2010) Personal email communication, 25th October.

Lühr, R. (2010) ‚Indogermanisch-Germanisch-Deutsch', in Schmid, H. U. & Ziegler, A. (eds.) *Jahrbuch für Germanistische Sprachgeschichte*. Berlin: de Gruyter, 163-187.

Maas, H-D. (1998) ‚Multilinguale Textverarbeitung mit MPRO' in *Europäische Kommunikationskybernetik heute und morgen*. Paderborn.

Mackey, A. & Gass, S.M. (2012) Research methods in Second Language Acquisition: a practical guide. Chichester: Wiley Blackwell.

Mackey, A., Abbuhl, R. & Gass, S.M. (2012) 'Interactionist Approach' in Gass, S. M. & Mackey, A. (eds.) The Routledge Handbook of Second Language Acquisition. London, Routledge, 7-23.

MacWhinney, B. (2000) *The CHILDES Project: Tools for Analyzing Talk.* 3rd edn. Mahwah, NJ: Lawrence Erlbaum Associates.

Marcus, M. P., Santorini, B. & Marcinkiewicz, M. A. (1993) 'Building a large annotated corpus of English: the Penn Treebank'. *Computational Linguistics*, 19, 313-330.

Mills, A. E. (1985) 'The acquisition of German' in Slobin, D. I. (ed.) *The Crosslinguistic Study of Language Acquisition* (vol. 1). Mahwah, NJ: Lawrence Erlbaum, 141-254.

Moskovsky, C. & Libert, A. (2004) 'Possibilities for passives in natural and artificial languages'. *Journal of Universal Language*, 5, 101-149.

Myles, F. (2004) 'From data to theory: the over-representation of linguistic knowledge in SLA'. *Transactions of the Philological Society*, 102, 139-168.

Myles, F. (2005) 'Interlanguage corpora and SLA research'. *Second Language Research*, 21, 373-391.

Norris, J. M. & Ortega, L. (2001) 'Effectiveness of L2 instruction: a research synthesis and quantitative meta-analysis'. *Language Learning*, 50, 417-528.

Norris, J. M. & Ortega, L. (2003) 'Defining and measuring SLA' in Doughty, C. J. & Long, M. H. (eds.) *The Handbook of Second Language Acquisition*. Oxford: Blackwell, 716-761.

Norris, J. M. & Ortega, L. (2012) 'Assessing learner knowledge' in Gass, S. M. & Mackey, A. (eds.) *The Routledge Handbook of Second Language Acquisition*. London, Routledge, 573-589.

Ortega, L. (2003) 'Syntactic complexity measures and their relationship to L2 proficiency: a research synthesis of college-level L2 writing'. *Applied Linguistics*, 24, 492-518.

Pallotti, G. (2007) 'An operational definition of the emergence criterion'. *Applied Linguistics*, 28, 361-382.

Papp, S. (2000) 'Stable and developmental optionality in native and non-native Hungarian grammars'. *Second Language Research*, 16, 173-200.

Pfenninger, S. (2013) 'Quadrilingual advantages: *do*-support in bilingual vs. multilingual learners'. *International Journal of Multilingualism*, 11, 143-163.

Pica, T. (1983) 'Methods of morpheme quantification: their effect on the interpretation of second language data'. *Studies in Second Language Acquisition,* 6, 69-78.

Pienemann, M. (1980) 'The second language acquisition of immigrant children', in Felix, S. (ed.) *Second Language Development: Trends and Issues.* Tübingen: Narr.

Pienemann, M. (1989) 'Is language teachable? Psychological experiments and hypotheses'. *Applied Linguistics*, 10, 52-79.

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

Pinker, S. (1996) *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.

Pinker, S., Lebeaux, D. S. & Frost, L. A. (1987) 'Productivity and constraints in the acquisition of the passive'. *Cognition*, 26, 195-267.

Polio, C. G. (1997) 'Measures of linguistic accuracy in second language writing research'. *Language Learning*, 47, 101-143.

Pravec, N. A. (2002) 'Survey of learner corpora'. *ICAME Journal*, 26, 81-114.

Rankin, T. (2011) 'The transfer of V2: inversion and negation in German and Dutch learners of English'. International Journal of Bilingualism, 16, 139-158.

Rastelli, S. (2009) 'Learner corpora without error tagging'. Available at: http://www.linguistik-online.de/38_09/rastelli.html. (Accessed 9.10.2010.)

Rayson, P. & Garside, R. (2000) 'Comparing corpora using frequency profiling' in *Proceedings of the Workshop on Comparing Corpora* held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics (ACL 2000), 1-8 October 2000, Hong Kong, 1-6.

Richards, B. (1987) 'Type-token ratios: what do they really tell us?' *Journal of Child Language,* 14, 201-209.

Roberts, L. (2013) IRIS seminar, University of York, 2nd-3rd September. Personal communication.

Schmid, H. & Laws, F. (2008) 'Estimation of conditional probabilities with decision trees and an application to fine-grained POS tagging'. *COLING 2008*. Manchester 18-22 August. Stroudsburg, Pa: Association for Computational Linguistics, 777-784.

Schwartz, B. & Sprouse, R. A. (1994) 'Word order and nominative case in non-native language acquisition. A longitudinal study of (L1 Turkish) German interlanguage' in Hoekstra, T. & Schwartz, B. (eds.) *Language Acquisition Studies in Generative Grammar*. Amsterdam: John Benjamins, 317-368.

Schwartz, B. & Sprouse, R. A. (1996) 'L2 cognitive states and the full transfer/full access model'. *Second Language Research*, 12, 40-72.

Scott, M. (2004) WordSmith Tools. [4] Oxford: Oxford University Press. [Computer Program].

Scott, M. (2010) 'What can corpus software do?', in O'Keefe, A. & McCarthy, M. (eds.) *The Routledge Handbook of Corpus Linguistics*. London: Routledge, 136-152.

Selinker, L. (1972) 'Interlanguage'. International Review of Applied Linguistics in Language Teaching, 10, 209-231.

Sharwood Smith, M. A. (1981) 'Consciousness-raising and the second language learner'. *Applied Linguistics*, 2, 159-168.

Sharwood Smith, M. A. (1993) 'Input enhancement in instructed SLA'. *Studies in Second Language Acquisition*, 15, 165-179.

Sharwood Smith, M.A. (1996) 'Cross-linguistic influence with special reference to the acquisition of grammar', in Jordens, P. & Lallemans J.A. (eds.) *Investigating Second Language Acquisition*. Berlin: Mouton de Gruyter, 71-86.

Sharwood Smith, M. A. (2013) Personal communication, July 30th.

Sharwood Smith, M. A. & Truscott, J. (2005) 'Stages or continua in second language acquisition: A MOGUL solution'. *Applied Linguistics*, 26, 219-240.

Sharwood Smith, M., Truscott, J. & Hawkins, R. (2013) 'Explaining change in transition grammars' in Herschensohn, J. & Young-Scholten, M. (eds.) *The Cambridge Handbook of Second Language Acquisition.* Cambridge, Cambridge University Press, 560-580.

Sinclair, J. (2004) 'Intuition and annotation – the discussion continues', in Aijmer, K. & Altenberg, B. (eds.), *Advances In Corpus Linguistics*, Proceedings of the 23rd International Conference on English Language Research on Computer Corpora (ICAME 23), 39-59.

Slabakova, R. (2013) 'Discourse and pragmatics' in Herschensohn, J. & Young-Scholten, M. (eds.) *The Cambridge Handbook of Second Language Acquisition.* Cambridge, Cambridge University Press, 482-584.

Slobin, D. I. (1985) *Crosslinguistic study of language acquisition* (vol. 1). Mahwah, NJ: Lawrence Erlbaum Associates.

Slobin, D. I. (1996) "From 'thought and language' to 'thinking for speaking'", in Gumperz, J. J. & Levinson, S. C. (eds.) *Rethinking linguistic relativity. Studies in the Social and Cultural Foundations of Language* 1. Cambridge: Cambridge University Press, 70-96.

Smith, N., Hoffmann, S. & Rayson, P. (2008) 'Corpus Tools and Methods, Today and Tomorrow: Incorporating Linguists' Manual Annotations'. *Literary and Linguistic Computing,* 23, 162-180.

Sorace, A. (2000) 'Syntactic optionality in non-native grammars'. *Second Language Research*, 16, 93-102.

Sorace, A. & Filiaci, F. (2006) 'Anaphora resolution in near-native speakers of Italian'. *Second Language Research*, 22, 339-368.
Available at: http://search.proquest.com/docview/814196111. (Accessed 27.07.2013.)

Spada, N. (2007) 'Communicative Language teaching: Current Status and Future Prospects', in Cummins, J. & Davison, C. (eds.) *International Handbook of English Language Teaching*. New York, Springer, 271-288.

St. John, M. F. & Gernsbacher, M. A. (1998) 'Learning and losing syntax: practice makes perfect and frequency builds fortitude', in Healy, A. F. & Bourne, L. E. (eds.) *Foreign Language Learning: Psycholinguistic Studies on Training and Retention*. Mahwah, NJ: Lawrence Erlbaum Associates, 231-255.

Stake, R. E. (2000) 'Case studies', in Denzin, N. K. & Lincoln, Y. S. (eds.) *Handbook of Qualitative Research*. Thousand Oaks, CA, London & New Delhi: Sage, 435-454.

Szagun, G. (2002) 'Learning the h(e)ard way: The acquisition of grammar in young German speaking children with cochlear implants and with normal hearing', in Windsor, F., Kelly, M. L. & Hewlett, N. (eds.) *Investigations in Clinical Phonetics and Linguistics*, Mahwah, NJ: Lawrence Erlbaum Associates, 131-144.

Thoday, E. (2008) 'Issues in building learner corpora: an investigation into the acquisition of German passive constructions' in *Newcastle Working Papers in Linguistics*, 14, 145-155.

Université Catholique de Louvain (2013) Centre for English Corpus Linguistics. Available at: http://www.uclouvain.be/en-cecl-lcWorld.html. (Accessed 20.12.2013.)

Vainikka, A. & Young-Scholten, M. (1994) 'Direct access to X'-theory: evidence from Korean and Turkish adults learning German', in Hoekstra, T. & Schwartz, B. D. (eds.) *Language Acquisition in Generative Grammar.* Amsterdam: John Benjamins, 265-316.

VanPatten, B. & Williams, J. (2007) *Theories in Second Language Acquisition,* Mahwah, NJ: Lawrence Erlbaum Associates.

Vogel, P. M. (2003) ‚Passiv in deutschsprachigen Chats. Eine Korpusanalyse'. Available at http://www.linguistik-online.de/15_03/vogel.html. (Accessed 6.10.2013.)

Wang, Y. (2010) 'Classification and SLA studies of passive voice'. *Journal of Language Teaching and Research*, 1, 945-949.

Weinberg, A. (1978) 'Comments on Borer and Wexler', in Roeper, T. & Williams, E. (eds.) *Parameter Setting*. Dordrecht: D. Reidel Publishing Company, 173-187.

Weinberger, U. (2002) Error analysis with computer learner corpora. A corpus-based study of errors in the written German of British university students. MA thesis, University of Lancaster.

White, L. (2003) 'On the nature of interlanguage representation', in Doughty, J. & Long, M. H. (eds.) *The Handbook of Second Language Acquisition*. Oxford: Blackwell, 19-42.

Wray, A. (2002) *Formulaic Language and the Lexicon.* Cambridge: Cambridge University Press.

Wray, A. (2007) 'Set phrases in second language acquisition' in Burger, H., Dobrovol'skij, D., Kühn, P. &Norrick, N. (eds.) *Phraseologie. Ein internationales Handbuch der zeitgenössischen Forschung* (vol.1)*.* Berlin: de Gruyter, 870-881.

Wray, A. (2008) *Formulaic Language: Pushing the Boundaries.* Oxford: Oxford University Press.

Wray, A. (2013) 'Formulaic language'. *Language Teaching*, 46, 316-334.

Zeldes, A., Ritz, J., Lüdeling, A. & Chiarcos, C. (2009) ANNIS: a search tool for multi-layer annotated corpora. Available at:
 http://zope.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/mitarbeiter-innen/amir/pdf/CL2009_ANNIS_pre.pdf (Accessed 5.12.2013.)

Zorach, C. & Melin, C. (1994). *English Grammar for Students of German*. Ann Arbor, MI: Olivia and Hill Press.