

Overt and Null
Subjects in Bulgarian
and in L1 Bulgarian-
L2 German
Interlanguage

Overt and Null Subjects in Bulgarian and in L1 Bulgarian- L2 German Interlanguage

By

Dobrinka Genevska-Hanke

Cambridge
Scholars
Publishing



Overt and Null Subjects in Bulgarian and in L1 Bulgarian-L2 German
Interlanguage

By Dobrinka Genevska-Hanke

Dissertation from the Carl von Ossietzky university of Oldenburg,
Oldenburg, Germany

This book first published 2019

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Copyright © 2019 by Dobrinka Genevska-Hanke

All rights for this book reserved. No part of this book may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
electronic, mechanical, photocopying, recording or otherwise, without
the prior permission of the copyright owner.

ISBN (10): 1-5275-2353-5

ISBN (13): 978-1-5275-2353-1

To my dearest ones, in deep gratitude for all the support that made this
work possible

*Аз съм гладна съм.
az sam gladna sam
I am hungry am *pro*.
'I am hungry.'

Larissa 5;5

TABLE OF CONTENTS

List of Illustrations	ix
List of Tables	xii
Acknowledgements	xiii
Chapter One.....	1
General Introduction	
1.1 Introducing Bulgarian Null Subjects.....	1
1.2 Subjects in L2 Development: L1 Bulgarian-L2 German	7
1.3 Aims, Background and Outline of the Present Work.....	8
Chapter Two	16
Theoretical Background on Null Subjects	
Introduction.....	16
2.1 The Null Subject Parameter	17
2.2 Null Subjects Cross-Linguistically	37
2.3 Problems of Classification	47
2.4 A Typology of Pronouns.....	61
2.5 Information Structure Effects on the Syntax of Subjects	64
Summary.....	72
Chapter Three	74
Empirical Investigations of Bulgarian Subjects	
Introduction.....	74
3.1 Descriptive Properties of Subjects and Conditions of Their Use... ..	75
3.2 Null Subjects in Spontaneous Speech: The Corpus	86
3.3 A Historical Case Study: Null Subjects in the 19 th century	95
3.4 Elicited Production of New Information Focus: Postverbal Subjects?	100
3.5 General Discussion and Implications.....	115
Summary.....	131

Chapter Four.....	134
L2 Acquisition Theory	
Introduction.....	134
4.1 First and Second/Foreign Language Acquisition in Contrast.....	135
4.2 Defining Interlanguage Grammars.....	145
4.3 Concepts in Generative L2 Theory	149
Summary.....	159
Chapter Five	161
A Comparative Background: Bulgarian and German Subjects	
Introduction.....	161
5.1 Introducing German Subjects.....	161
5.2 Conflicting Aspects as to Subject Use	168
Summary.....	170
Chapter Six.....	171
Empirical Investigations of L1 Bulgarian-L2 German Interlanguage	
Introduction.....	171
6.1 Participants and Procedure.....	171
6.2 Aims and Predictions	181
6.3 Overall Results and Discussion.....	183
6.4 Results per Condition.....	193
6.5 General Discussion: How Native is Near-native?.....	224
Summary.....	228
Chapter Seven.....	230
Conclusion and Outlook	
References	237
Appendices	263
Appendix A Questionnaire on Language Background	
Appendix B Grammaticality Judgment and Interpretation Task	
Appendix C GJT/IT – Items per Condition Listing	
Appendix D Questions of the Video Task on New Information Focus	
Index.....	295

LIST OF ILLUSTRATIONS

Figure 3-1 Overall rates for types of subject IFoc in Bulgarian in percentages	106
Figure 3-2 Overall rates for types of subject IFoc for the region of Varna in percentages.....	108
Figure 3-3 Overall rates for types of subject IFoc for the region of Sofia in percentages	108
Figure 3-4 Interregional difference postverbal subjects (all generations) – distribution in percentages	109
Figure 3-5 Postverbal subjects for generation 1 across regions – difference in percentages	110
Figure 3-6 Postverbal subjects for generation 2 across regions – difference in percentages	110
Figure 3-7 Intergenerational difference postverbal subjects Varna – distribution in percentages	112
Figure 3-8 Rates in percentages (all answering strategies) for the region of Varna split to generation	112
Figure 3-9 Intergenerational difference postverbal subjects Sofia – distribution in percentages	113
Figure 3-10 Rates in percentages (all answering strategies) for the region of Sofia split to generation.....	114
Figure 3-11 Rates for object IFoc types for the region of Varna in percentages	114
Figure 3-12 Italian subject IFoc types	128
Figure 3-13 BP subject IFoc types.....	129
Figure 3-14 Finnish subject IFoc types	129
Figure 6-1 GJT task description and familiarizing examples	174
Figure 6-2a,b Mean group rates per subcondition of the GJT on subject use - German vs. Bulgarian.....	190
Figure 6-3 Within-group distribution of GJT rates on subject use in Expletive Subordinate Clause (SC) – single Bulgarian group	194
Figure 6-4 Within-group distribution of GJT rates on subject use in Expletive Subordinate Clause (SC) – Bulgarian HC vs. TC.....	194

Figure 6-5 Within-group distribution of GJT rates on subject use in Expletive Subordinate Clause (SC) – Bulgarian HC vs. TCSR vs. TCLR	195
Figure 6-6 Expletive Subordinate Clause (SC) NS – Incorrectness scores of subject use in percentages.....	197
Figure 6-7 Expletive Subordinate Clause (SC) OS – Incorrectness scores of subject use in percentages.....	197
Figure 6-8 Within-group distribution of GJT rates on subject use in Expletive Main Clause (MC) – single Bulgarian group.....	201
Figure 6-9 Within-group distribution of GJT rates on subject use in Expletive Main Clause (MC) – Bulgarian HC vs. TC	201
Figure 6-10 Within-group distribution of GJT rates on subject use in Expletive Main Clause (MC) – Bulgarian HC vs. TCSR vs. TCLR..	202
Figure 6-11 Within-group distribution of GJT rates on subject use in Referential Clause-Initial (CI) – single Bulgarian group	203
Figure 6-12 Within-group distribution of GJT rates on subject use in Referential Clause-Initial (CI) – Bulgarian HC vs. TC.....	203
Figure 6-13 Within-group distribution of GJT rates on subject use in Referential Clause-Initial (CI) – Bulgarian HC vs. TCSR vs. TCLR...	204
Figure 6-14 Within-group distribution of GJT rates on subject use in Referential Clause-Medial (CM) – single Bulgarian group	205
Figure 6-15 Within-group distribution of GJT rates on subject use in Referential Clause-Medial (CM) – Bulgarian HC vs. TC.....	206
Figure 6-16 Within-group distribution of GJT rates on subject use in Referential Clause-Medial (CM) – Bulgarian HC vs. TCSR vs. TCLR	207
Figure 6-17 Referential Clause-Medial (CM) OS – Incorrectness scores of subject use in percentages.....	208
Figure 6-18 Referential Clause-Medial (CM) NS – Incorrectness scores of subject use in percentages.....	209
Figure 6-19 Within-group distribution of GJT rates on subject use in Coordination Same Verb (COOR SV) – single Bulgarian group	212
Figure 6-20 Within-group distribution of GJT rates on subject use in Coordination Same Verb (COOR SV) – Bulgarian HC vs. TC	213
Figure 6-21 Within-group distribution of GJT rates on subject use in Coordination Same Verb (COOR SV) – Bulgarian HC vs. TCSR vs. TCLR.....	213
Figure 6-22 Within-group distribution of GJT rates on subject use in Coordination Different Verb (COOR DV) – single Bulgarian group...	215
Figure 6-23 Within-group distribution of GJT rates on subject use in Coordination Different Verb (COOR DV) – Bulgarian HC vs. TC...	215

Figure 6-24 Within-group distribution of GJT rates on subject use in Coordination Different Verb (COOR DV) – Bulgarian HC vs. TCSR vs. TCLR.....	216
Figure 6-25 Coordination Different Verb (COOR DV) OS – Incorrectness scores of subject use in percentages.....	218
Figure 6-26 Coordination Different Verb (COOR DV) NS – Incorrectness scores of subject use in percentages.....	219
Figure 6-27 Within-group distribution of GJT rates on subject use in the V2 condition – Bulgarian HC vs. TC.....	220
Figure 6-28 V2 condition – Incorrectness scores of subject use in percentages	221
Figure 6-29 Within-group distribution of GJT rates in DP COOR Control – Bulgarian HC vs. TC	222

LIST OF TABLES

Table 2-1 Endings of Old Bulgarian past (imperfect) tense	28
Table 2-2 Unrelated classifications (discussed so far).....	58
Table 2-3 Some properties and their distribution in PNSLs and CNSLs...	60
Table 2-4 Typology of structural deficiency	61
Table 3-1 Bulgarian paradigms of personal pronouns.....	83
Table 3-2 Bulgarian sentence/subjects types	84
Table 3-3 Number and type of subjects in contemporary Bulgarian	89
Table 3-4 Distribution of subjects to clausal type	89
Table 3-5 Subjects split to different types	90
Table 3-6 Number and type of subjects in Bulgarian – contemporary and historical rates	98
Table 3-7 Cross-linguistic patterns of IFoc answering strategies	100
Table 3-8 Pro-drop rates cross-linguistically.....	116
Table 3-9 Typology of structural deficiency (extended for Bulgarian)...	125
Table 3-10 Answering strategies cross-linguistically – results of the video task	129
Table 5-1 Comparison of subject types and subject contexts – Bulgarian vs. German	169
Table 6-1 GJT participants	173
Table 6-2 Overall GJT results on subject use all groups (raw scores).....	185
Table 6-3 Interpretation Task – Overall results	223

ACKNOWLEDGEMENT

I would like to thank the reviewers for their work as well as my colleagues and supervisors for their insightful comments at various colloquia at the University of Oldenburg, especially Cornelia Hamann, Adriana Belletti and Gerd Hentschel. I'm particularly indebted to Istvan Fekete for his support with regard to the statistical analyses, to Teodora Todorova for her help with some of the data collection and transcription, to Margarita and Iva Mitevi for their help with some of the arrangements for the testings in Bulgaria and to Lauren Freede for reading the manuscript. I'm also grateful to the linguists at the workshops and conferences at which this work was presented for the discussions they engaged in. This project would not have been possible without the speakers tested in Germany and in Bulgaria, which I very much appreciated – thank you, I will never forget your participation. Special thanks go to one of the directors of the 5th Gymnasium for foreign languages *Ioan Ekzarkh* in Varna, Emilia Eneva, and to her students specializing in German who participated in this project in September 2011. The choice of this elite high school was not at all arbitrary since I myself graduated from it in 1999. The present work was in parts inspired by Cornelia Hamann, who is an expert in syntax and an excellent researcher and who patiently spent numerous hours on related discussion, valuable advice and encouragement – I cannot imagine a better supervisor than her. I can hardly express my gratitude for her being such an extraordinary mentor and a dear friend all the years. Without her, this work would not exist. Last but not least, I would like to thank my family and my friends for their incredible support along the way. Another person without whom this project would have never been accomplished is my husband Kristof – I can never really make up for it. And I thank my daughter Larissa for helping as much as she could, especially in taking care of her little sister Klara.

CHAPTER ONE

GENERAL INTRODUCTION

1.1 Introducing Bulgarian null subjects

Meine Eltern untereinander sprachen deutsch, wovon ich nichts verstehen durfte. Zu uns Kindern und zu allen Verwandten und Freunden sprachen sie spanisch... Die Bauernmädchen zuhause konnten nur Bulgarisch, und hauptsächlich mit ihnen wohl habe ich es auch gelernt.
(Canetti 2005, 17)

This quotation is from the introductory volume of Elias Canetti's biography. The author spent his early childhood in the Bulgarian city Russe, in the ghetto of the Spanish Jews. His family moved to Manchester when he entered primary school and later to Vienna, Zurich and Frankfurt, where Canetti spent his teenage years. He received the Nobel Prize for literature in 1981 and spent most of his lifetime in London and Zurich. What is particularly remarkable about the quotation given above is his acquisition of multiple languages in early childhood, Spanish, German and Bulgarian since multilingualism was much more limited at that time in Europe than it is today.

Language acquisition happens in different contexts and the present work is motivated by my own experience as a native Bulgarian who learned German and English as adult (in the sense of *late* or *post-puberty*) second/foreign languages. Additionally, two of the languages Canetti learnt as a child, Spanish and Bulgarian, are in typological opposition to German and English with regard to a prominent syntactic phenomenon – what is referred to as the *Null Subject Parameter*¹, which captures a major cross-linguistic difference. The present work focuses primarily on this

¹ The Null Subject Parameter is also known as the *Pro-drop Parameter* (both terms will be used interchangeably throughout the work). According to the *Extended Projection Principle* (EPP) of the Government and Binding Theory, null subjects are empty categories, lacking phonological content (and sometimes semantic content in the case of expletives), but present at all syntactic levels of representation (Chomsky 1981, 1982).

grammatical phenomenon as a possible source of problems in second language (L2) development. Its main aim is to research the L2 competence of highly proficient/near-native speakers of L2 German with Bulgarian as a first language (L1) in relation to the influence of the L1 on L2 subject use taking into consideration the typological difference mentioned above as well as language acquisition theory.

The term *parameter* captures the notion that languages differ as to certain grammatical properties. It is used within the generative framework, where parameters are considered the locus of cross-linguistic variation. The different settings of a parameter result from differences in the functional categories or their features across languages. The Null Subject Parameter accounts for cross-linguistic differences between languages with respect to subject use and the language difference in (1) in particular (e.g. Chomsky 1981, 1982; Rizzi 1982, 1986; Huang 1984, 1989, 2012; Jaeggli and Safir 1989; Holmberg 2005; Frascarelli 2007; Roberts and Holmberg 2010; Holmberg and Roberts 2013; van Gelderen 2013; Kayne 2013; but see also Sigurðsson 2011). The Bulgarian subjectless subordinate clause given in (1) “че *pro* идва” is grammatical, while its English (and German) counterparts * *that* \emptyset *comes* (* *dass* \emptyset *kommt*) are not.² In languages like Bulgarian, Italian, Spanish and Greek, recently classified as *consistent null subject languages*³ and traditionally known as *pro-drop languages*, null referential definite pronominal subjects, called *pro*, are licensed in finite clauses. The alternative with the overt pronominal subject given in (2) represents the only grammatical option in languages like English and German. (1) and (2) share one and the same reference in terms of each subject and the two subjects (the 1st person singular subject of the main clause and the 3rd person singular subject of the subordinate clause) are both referential.

² I use the symbols *pro* and \emptyset to indicate that an element corresponding to a grammatical subject (in the sense that it agrees with the verb) in tensed/finite clauses is missing. In contrast to *pro* which refers to an empty category, associated with the properties listed above, \emptyset does not refer to a particular type of empty category (empty categories are defined as they are typically defined in generative tradition). In some works the symbols *ec* or simply “—” are used instead. The type of empty element in the examples given throughout the text is represented as it is displayed in the original works from which they have been quoted, unless the examples are my own. Related comments will be supplied where necessary. Note that the term *subject* is used in a grammatical sense (see Keenan 1976; for subject-defining criteria). See Algeo (1966) for the transliteration from Cyrillic.

³ Note that the theoretical assumptions on null subjects and the related classification of languages have undergone considerable changes and that the term *null subject* is not restricted to null subjects of the type *pro* but is also used for all types of null subjects.

- | | |
|---|---|
| (1) <i>pro</i> знам, че <i>pro</i> идва.
<i>pro</i> znam, che <i>pro</i> idva
*Weiß, dass ∅ kommt
‘Ich weiß, dass er kommt.’
*Know that ∅ comes
‘I know that he’s coming.’ | (2) Аз знам, че той идва.
az znam, che toi idva
Ich weiß, dass er kommt
‘Ich weiß, dass er kommt.’
I know that he comes
‘I know that he’s coming.’ |
|---|---|

Pro-drop languages are languages in which the Null Subject Parameter is set positively. In contrast, in languages like English and German the parameter is set negatively, hence the ungrammaticality of (1).⁴ Grammatical person/number is encoded in the verb, which makes subject realization unnecessary. Rich verbal inflection in this type of languages is thus central for licensing null subjects. The possibility of having *pro* subjects in a language has related grammatical consequences – expletive null subjects, postverbal subjects and subject extraction from complement clauses are possible too (Rizzi 1982).

From a cross-linguistic point of view, pronominal subjects can generally be either overt or null. However, the choice between these two types is not solely motivated by the syntactic setting of the Null Subject Parameter. Other aspects of language such as information structure also play a role in governing their occurrence (Lambrecht 1994; Gundel and Fretheim 2004; Krifka 2006). In particular, constraints of information structure reach into the domain of syntax and determine subject choice or in other words, they are grammatically encoded (Rizzi 1997, 2004; Kiss 1998; Belletti 2004; Frascarelli and Hinterhölzl 2007; Chomsky 2008; Cruschina 2009; Aboh 2010). Some researchers propose that they originate in the grammar itself – the Complementizer phrase, CP or C-domain is the syntactic level aligning with discourse/pragmatics since it contains the functional categories Topic phrase (TopP) for topics and Focus phrase (FocP) for foci (Rizzi 1997). In a similar vein, the domain above the verbal phrase, the vP periphery, contains further projections of this kind (Belletti 2001 and subsequent work). The two notions *topic* and *focus* considerably affect syntax, be it through the presence or absence of subjects or different word order patterns matching distinct interpretations. One example is the use of a null subject in the context of *topic maintenance/continuity*, which is preferred by speakers of pro-drop languages as well as the obligatory insertion of an overt subject in the case of *topic shift*. The two sentences (3) and (4) have an identical

⁴ Null subjects in main clauses of spoken German like ∅ *kommt* are null topics, licensed through topic-drop (their reference is recovered through discourse, see Huang 1984; Cardinaletti 1990; Hamann 1996; Rizzi 2005a; Sigurdsson 2011; van Gelderen 2013).

translation in English and German but two different interpretations due to the distinct reference of the null and the overt pronoun:⁵

(3) Il professore_i ha parlato dopo che lui^{*i/j} è arrivato.
 the professor has spoken after he is arrived
 ‘The professor started speaking after he arrived.’

(4) Il professore_i ha parlato dopo che *pro*_{i/*j} è arrivato.
 the professor has spoken after is arrived
 ‘The professor started speaking after he arrived.’

This contrasts strongly with the state of affairs in non-pro-drop languages like English and German, where overt subjects appear in both cases, although there are differences as to the use of personal pronouns and nouns. In these languages, the overt pronoun is ambiguous in its reference. In pro-drop languages, the overt pronoun must differ in its reference from the subject of the main clause, while the null pronoun is identical to it (Samek-Ludovici 1996; Sorace 2005; Frascarelli 2007). In addition, focus plays an important role when it comes to word order and it makes the subject occur postverbally in new information contexts in some pro-drop languages (Belletti 2001 and subsequent work). The answer to a question like *Who spoke?* in Italian is necessarily a clause with a postverbal subject of the type illustrated in (5):

(5) Ha parlato Gianni.
 has spoken Gianni
 ‘Gianni spoke.’

According to Belletti’s analysis, the answer in English cannot include a postverbal subject unless the clause is a cleft and the typical answer to the question is an SV clause. These different types of answers are referred to as *answering strategies*.⁶ From a cross-linguistic point of view, for each language there is a predominant pattern as to the answering strategy employed – which can be linked to properties of focus expression. Because of the interaction of different domains, here syntax and pragmatics, such language phenomena have been labeled *interface phenomena*. In L2 settings (and according to the *Interface Hypothesis* proposed by Sorace 2005 and

⁵ The Italian examples (3) and (4) are from Roberts and Holmberg (2010). This pattern of null vs. overt subject alternation is also known as the *Avoid Pronoun Principle* in the generative framework.

⁶ *Strategy* refers to formal options which are constrained grammatically and pragmatically (Belletti 2008).

related work), the *external interface* between syntax on the one hand and discourse or information structure on the other hand has been seen as a source of misinterpretation on the side of the learner.

All the contrasts listed above create contexts for which the two languages of a speaker, the L1 and the L2, if one of them is a pro-drop language and the other one is not, make contradictory predictions (in grammatical terms). These contexts have been widely researched in the generative framework so that in the following, much related theoretical and empirical work will be reviewed and adopted. The contexts in question are additionally vulnerable in L2 development, because they cause learner errors or result in different preference patterns when it comes to language performance. Some researchers assume that learner grammars, called *interlanguage grammars*⁷ are permanently deficient, because problems of this kind are found to persist even at end stages of L2 development and after prolonged exposure to the language (e.g. Hawkins 2005; Tsimpli and Dimitrakoupoulou 2007; Prentza and Tsimpli 2013; Di Domenico 2015). They have been further attributed to the incomplete acquisition of functional categories or features thereof as well as to the way these are mapped/assembled in a language, since this affects the instantiation of different parameter settings and gives rise to diverging word orders cross-linguistically. For instance, the functional category FocP might be present (active) as in Italian or missing and thus inactive as in English, where *in situ* focalization with specific prosodic marking is applied instead (Belletti 2007). Accordingly, the word order pattern with a postverbal subject illustrated in (5) is possible in Italian but not in English. This is connected to the types of subjects available in a given language since different types of subjects have distinct properties, giving rise to a certain featural configuration or *assembly* for a particular element across languages (Lardiere 2007). This raises the question of the learnability of the respective elements. Bley-Vroman (1990) proposes that the difference between L1 and L2 grammars is fundamental, an assumption related to the much-debated concept of a critical period for language, in which age of acquisition effects play a central role (Johnson and Newport 1989; Birdsong 1992, 1999, 2005; White and Genesee 1996; Bialystok and Hakuta 1999; De Keyser 2000; Hyltenstam and Abrahamsson 2000, 2003; DeKeyser and Larson-Hall 2005; Abrahamsson and Hyltenstam 2009). Some researchers claim that language is learnt from experience and that it is deeply affected by communication and social interaction instead (e.g. Lieven 1994; Plunkett and Elman 1997; Tomasello 2015).

⁷ The term goes back to Selinker (1969, 1972).

One of the key questions of L2 acquisition theory is whether the underlying knowledge representations of L2 learners who can be hardly distinguished from native speakers are native-like or not, and if L2 and native speakers use the same cognitive resources and processing strategies; for some researchers differences of this kind exist but others attribute apparent similarities to other factors of influence (Birdsong 1992, 1999, 2005; Ioup et al. 1994; White and Genesee 1996; Hawkins and Chan 1997; Liceras 1998; Hudson and Newport 1999; Hyltenstam and Abrahamsson 2003; Hawkins 2005; De Keyser and Larson-Hall 2005; Clahsen and Felser 2006; Meisel 2009; Felser and Clahsen 2009). In addition, some researchers suggest that L2 processing is guided by semantic-conceptual analyses, because particular L2 functional categories or features thereof are missing in the learner competence (e.g. Clahsen and Felser 2006). Within generative theory, there are what are known as the representational and processing accounts. The former assume grammatical deficits in interlanguage competence and seek the answer to a seemingly native-like performance in avoidance strategies and formulaic knowledge on the side of the learner. In contrast, processing accounts attribute L2 problems to deficient performance. Two hypotheses have been particularly influential in recent research on pronominal use – the *Interface Hypothesis* and the *Interpretability Hypothesis* and both will be taken into consideration for the L2 study of the present work. Needless to say, researchers need to look for individuals who have reached the end state of their L2 grammar in order to answer these questions in a satisfactory manner.

As to the L2 research available on the Null Subject Parameter, there is a number of theoretical and/or empirical studies, mainly testing the performance of speakers with a non-pro-drop L1 (typically English) who learn a pro-drop language. One of the pioneering studies is by Liceras (1988) who tested L2 Spanish. Some of the more recent studies are those by Sorace (2003), Belletti, Bennati and Sorace (2007) and Di Domenico (2015) for L2 Italian, Hertel (2003), Montrul and Rodríguez-Louro (2006), Ballester (2007) and Dominguez and Arche (2008) for L2 Spanish, Tsimpli and Sorace (2006) for L2 Greek. See e.g. Serratrice (2007), Dal Pozzo (2011) and Pinto (2011) for studies carried out with bilinguals. There are also studies reporting on learners of a non-pro-drop L2 with a pro-drop L1. For instance, Tsimpli and Roussou (1991), Prentza and Tsimpli (2013) and Prentza (2014) show that subject use of L2 speakers of English with L1 Greek significantly differs from that of native speakers. The results of Tsimpli and Roussou (1991) revealed L1 transfer of the positive value of the Pro-drop Parameter even for proficient L2 speakers, which made the learners analyze overt subjects as agreement elements. The authors interpret

this finding as evidence for a failure to reset parameters. In the same spirit, White (1985) tested native speakers of Spanish in their L2 English with the result that they used up to 40% of null subjects in English. A more recent study by Prentza and Tsimpli (2013) shows that L2 speakers of English use ungrammatical null subjects in adjunct CP clauses (like *The student was upset because he had failed the test*) and prefer overt subjects in VP-coordination structures (like *Jane had studied hard and (she) passed the exam*) to a significantly greater extent than native speakers do. Similarly, Prentza (2014) shows long-lasting problems with structures involving null and postverbal subjects as well as that-trace violations in the L2 English of L1 Greek speakers with advanced L2 proficiency.⁸ Again, this points to problems with parameter resetting (see also Liceras 1989; Lozano 2002; and Lardiere 2007). A study by Sopata (2005) investigates the L2 competence in (non-pro-drop) German of speakers with (pro-drop) L1 Polish. Her results show residual deficits in terms of a deficient interlanguage grammar – the participants who were proficient in their L2 still had problems with German null expletive subjects.

1.2 Subjects in L2 development: L1 Bulgarian-L2 German

The typological difference under consideration in this work is one between the Slavic pro-drop language (or consistent null subject language) Bulgarian and the Germanic non-pro-drop language (or semi-null subject language) German, which allows null topics in its spoken register (see Bojadziev, Kuzarov, and Penchev 1999; and Popov 1998; for Bulgarian, Hamann 1996; Rizzi 2005a; Roberts and Holmberg 2010; Biberauer 2010; Barbosa 2011; Trutkowski 2011, 2016; for German). In addition to the null subjects illustrated in the examples so far, there are also two other types of null subjects, relevant for the current comparison, both of which are typical of German. One type is expletive null subjects, which are the only type of null subjects in a distinct group of null subject languages, called semi-null subject languages (Roberts and Holmberg 2010; Biberauer 2010; Barbosa 2011). German, Jamaican and some Dutch varieties are languages of this type. Expletives are non-referential and do not carry semantic meaning, see example (6).⁹

⁸ Note that using null subjects in adjunct CP clauses and subject extraction from that-clauses are ungrammatical in English but crucially, grammatical in Greek. The VP-coordination structure could be viewed as equally good in both languages and using an overt or a null subject is a matter of preference.

⁹ A referential reading is possible here but excluded of consideration.

- (6) Gestern wurde (*es) getanzt.
 yesterday was (*it) danced
 'People danced yesterday.'

Note that if the constituent *gestern* in the meaning of *yesterday* does not occupy the clause-initial position and the expletive *es* in the meaning of *it* surfaces clause-initially, it necessarily has to be overt. The other type of null subject which is possible in German is a null topic (Huang 1984, 1989; Cardinaletti 1990; Hamann 1996; Rizzi 2005a; Trutkowski 2011, 2016), compare (7).

- (7) ∅ hab(e) es gestern gekauft.
 ∅ have1SG it yesterday bought
 'I bought it yesterday.'

Null topics are referential null subjects, licensed through the grammatical mechanism of operator identification. This operates by discourse means and is typical of languages like Chinese.¹⁰ The licensing of null topics is traditionally viewed as different from that of the null subject in pro-drop languages.

As mentioned, interlanguage problems related to subject use have been attested not only in the learner grammars of speakers of a non-pro-drop L1 learning a pro-drop L2, but also for speakers of a pro-drop L1 who are learning a non-pro-drop L2. Related language deficits are harder to detect in the latter than in the former case, since they are less obvious. The combination of German as L2 and Bulgarian as a pro-drop L1 is a language constellation of this kind.

1.3 Aims, background and outline of the present work

Although numerous studies have been carried out in the domain of research on adult second and foreign language development during the past few decades, there are still many open questions. As mentioned, one of these is the question of the limits of ultimate attainment or in other words, whether interlanguage grammars of adult near-native L2 speakers can become indistinguishable from native grammars. A related question is that of the kind of underlying mental representations for these interlanguage grammars.

¹⁰ German is much more restrictive than Chinese in the use of topic-drop: details follow.

As a result, this study has implications for both linguistic theory and language teaching.

The goal of the present piece of original research, consisting of several empirical studies, is to contribute to answering these big questions by testing very proficient, near-native foreign and second language speakers of the language combination L1 Bulgarian-L2 German. The language combination of a Slavic consistent null subject (pro-drop) as L1 and a Germanic semi-null subject (non-pro-drop but expletive and topic-drop) as L2 has not been investigated before as to the domain of pronominal knowledge. This book therefore explores the realization of pronominal subjects in Bulgarian and its implications for adult near-native competence of German as a second/foreign language. Looking at Bulgarian is particularly interesting in this context, not simply because the language is not well-researched but because it seems to differ from other pro-drop languages, displaying features of non-pro-drop languages at the same time. For this reason, novel typological investigations on subject realization in spontaneous speech as well as an oral elicitation task on new information focus were carried out with three generations of monolingual Bulgarian speakers to serve as a background for the empirical study of near-native L2 subject use. Accordingly, the main aims of this piece of work were twofold – adequately classifying Bulgarian as to its pro-drop nature and determining the possible impact of related cross-linguistic differences on near-native interlanguage grammars of speakers with the language combination pro-drop L1 Bulgarian/non-pro-drop L2 German.

Importantly, although German is not pro-drop, it allows null topical subjects and requires some obligatory null expletive subjects, so that some null subject contexts superficially overlap for the two languages. This might cause interlanguage problems if no proper differentiation between subject types is made. In comparison to the wide use of null referential subjects in Bulgarian, in German there are two types of null subjects that are much more restricted. Since the language context is one of a semi-null subject language where overt subjects occur in the vast majority of cases, there is direct evidence for the prominence of overt subjects in the language input. Overt subjects are what L2 speakers experience most of the time in the language surrounding them and missing subjects in the language of L2 speakers are often corrected by native speakers and teachers. The general expectation is that the more proficient L1 Bulgarian speakers become in their L2 German, the fewer null subjects they use. However, it is an open question as to whether they have acquired the overt pronouns of their L2 in terms of the features and grammatical principles ruling their appearance or whether they instead apply a learner strategy of using an overt subject in

every clause, which would mask performance and override the optional null topics and the obligatory null expletives of German.¹¹ Since overt topics are optional and their use alongside null topics grammatical, I speak of a possibly different preference pattern for L2 speakers in the following. In contrast, the use of overt expletives in contexts in which they are prohibited is considered erroneous since it gives rise to strong ungrammaticality. In addition, the question will be pursued as to whether L2 speakers are able to fully distinguish different null subject types – the *pro* subjects of their native language as opposed to the null topics and null expletive subjects of their L2 German. This would suggest that the features of the L2 subjects are acquired or appropriately assembled by near-native adult learners.

The empirical data set of the present study consists of language tests as well as recordings of spontaneous speech production carried out with 101 native speakers of Bulgarian and 20 native speakers of German, accompanied by a questionnaire on language background. 51 of the Bulgarians are proficient speakers of German as a foreign language, 31 of whom are foreign language learners and 20 L2 learners in the target language country. The latter group of speakers had been German residents for a prolonged period of time, their L1 use very restricted and their exposure to German massive prior to and at the time of investigation. The German speakers are the control group. The language test conducted on German consists of a grammaticality judgment task and an interpretation task, including different conditions with topics and expletives as well as subjects in coordination structures (the second conjunct can be generally subjectless across languages). The idea behind this was to research the L2 patterns of null subject use, comparing language-dependent structures with structures that are not subject to cross-linguistic difference and thus universally available across languages, and to see whether non-native speakers are sensitive to the different overt and null subject contexts of their L2. The other 50 Bulgarian speakers were tested in their native language. A corpus of spontaneous speech production was created, consisting of the recordings of ten Bulgarian speakers (all Bulgarian residents, raised by Bulgarian monolinguals). All 50 speakers were tested on answering strategies so that the word order pattern of new information focus in the language could be determined. The speakers belong to three different generations and two geographical regions so that possible related influence could be accounted for. This was accomplished with the help of a video

¹¹ Note that rules of this kind are not explicitly formulated in L2 German textbooks. This idea is further in line with the related assumption on the analysis of overt subjects as agreement elements (Tsimpli 2007).

task, first developed and employed for Italian by Belletti and Leonini (2004) and thereafter used for multiple languages. This task was extended in order to test for the position of new information focus objects in addition to subjects. The Bulgarian data was gathered with the purpose of investigating the synchronic profile of Bulgarian with respect to subject use.¹² In addition, as these results revealed parallels to Brazilian Portuguese (BP), diachronic data was analyzed and compared as well. The three typological data sets provide a solid basis for related L1 facts and assumptions as well as for the present L2 study.

Comparative syntax is the focus throughout this work and theoretical frameworks are critically evaluated as to their explanatory power from a cross-linguistic perspective. In particular, Bulgarian cannot easily be fit into any of the language classes proposed in the recent typological classification of languages as to the availability of different kinds of null subjects by Roberts and Holmberg (2010), despite its description as a pro-drop language by Bojadziev, Kuzarov, and Penchev (1999). Importantly, Bulgarian shows the presence of several features that are typical of non-pro-drop languages but this does not seem to affect the positive setting of the Null Subject Parameter. Additionally problematic is the syncretism in its inflectional paradigm for past tense, because according to the proposal on pro-drop and impoverishment by Müller (2006) its presence should exclude pro-drop as a co-occurring property. The fact that most of these properties have also been attested for other pro-drop languages possibly points to a more general problem of classification and further challenges some of the assumptions of Roberts and Holmberg (2010), going back to Holmberg (2005) and related work. Accordingly, future research is necessary for a better understanding of the complex interplay of these properties with the null subject property, as well as for modifications of related typological classifications.

The novel results on subject realization in Bulgarian obtained here provide evidence that Bulgarian is a pro-drop language with overall high rates of null subjects, especially in subordinate clauses, which is a solid indication for pro-drop. The findings also show that postverbal subjects are untypical in new information focus contexts, another novel result of the typological investigation on Bulgarian. Overall, the pattern attested for Bulgarian was rather mixed and the postverbal subject position, yielding the word order verb subject (VS), was only exploited in up to 30% of the cases. This contrasts with the highly consistent pattern of pro-drop Italian (Belletti

¹² Recall that the availability of postverbal subjects in a language is related to the null subject property, so that specific answering strategies are typical of pro-drop languages. The test on answering strategies in focal contexts was thus run in addition to the analysis of spontaneous speech production data.

2009) and is similar to the patterns found in other partial/non-pro-drop languages. Since the video task has not been used for other pro-drop languages and since postverbal subjects are a property which is more loosely connected to the null subject property than previously thought, the findings are indicative for classification matters. The results further revealed regional and generational differences, suggestive of a possible syntactic change independent of the null subject property. Furthermore, a loss of null subjects as in BP seems improbable for Bulgarian since data from an earlier stage of the language was also analyzed and showed no differences in comparison to the contemporary data. A close inspection of the topic continuity contexts in which overt subjects occurred led to the assumption that these subjects are *weak* in nature in the sense of Cardinaletti and Starke (1994). This explains their ability to alternate with the null subject *pro* and in a subsequent step to the postulation of a weak pronominal paradigm for Bulgarian subjects. This is a novel third pronominal paradigm next to the established ones for strong subjects and clitics, formally identical to that of strong subjects. Importantly, overt subjects were mostly used by the speakers in the corpus because they occurred in grammatical contexts that require overt subjects. I conclude that as long as inflection is rich, Bulgarian will remain a pro-drop language. Accordingly, the evolution of a pronominal paradigm for weak subjects and the establishment of a word order pattern for focus which does not involve a null subject are insufficient to cause the loss of null subjects in Bulgarian, calling into question Robert's proposal on the loss of the null subject property (2014).

Turning our attention to theories of L2 development and looking at the Interface Hypothesis in particular, its main prediction is that of optionality for specific properties of language, namely features of the *interface syntax*, which might be deficient, while features of *core/narrow syntax* remain spared in both L2 development and L1 attrition. Furthermore, it has been suggested that cross-linguistic differences play a minor role, if any (e.g. Sorace et al. 2009). The latter has already been challenged by research showing that differences exist between Italian and Spanish as to the scope of overt pronouns (Filiaci 2010; see also Prentza and Tsimpli 2013; for similar differences between Greek and Spanish). Furthermore, the present L2 study provides solid evidence for permanent deficits in the domain of narrow syntax for L2, which is in line with the results of other similar studies. Accordingly, the Interface Hypothesis as a common framework of bilingualism, i.e. for L1 attrition as well as L2 development, possibly needs modification of its predictions for the latter, which in turn points to the fact that L2 development might be crucially different from L1 attrition. In contrast, the *Fundamental Difference Hypothesis*, the *Interpretability*

Hypothesis and the *Shallow Processing Hypothesis* (Clahsen and Felser 2006), among others, all predict permanent deficits in L2 knowledge representations and/or language processing for adult L2 speakers. The present L2 study provides evidence for these predictions, testing pronominal knowledge in L2 speakers of a language combination which, to my knowledge, has not been investigated before.

The results of the L2 study revealed that the performance of the L2 speakers differed significantly from that of the native German control group. No differences between the foreign and the second language learners were attested. The L2 speakers had a strong preference for overt subject use across contexts/conditions, including contexts in which these subjects are ungrammatical and contexts in which null subjects are possible language-independently (coordination structures), which is in line with some but contradicts other previous research. Furthermore, the findings additionally challenge theoretical proposals supporting native-like attainment in L2 development and the successful acquisition of purely grammatical features. At the same time, they support proposals on fundamental differences between first and second languages as well as proposals on permanent deficits in underlying L2 knowledge representations or in other words, L2 competence. Importantly, the L2 speakers of the present study were insensitive to subject type and used ungrammatical subjects while rejecting grammatical ones, which resulted in performance patterns significantly different from those of native speakers and thus indicative of interference with L1 grammatical features. The problems concerned both German null topics and null expletives. In particular, the near-native L2 speakers failed to differentiate between grammatical clause-initial and ungrammatical clause-medial null referential subjects/topics. In contrast, the German controls supplied the right types of subjects into the corresponding contexts for subjects. Furthermore, the near-native speakers used all possible subject types – referential, expletive and arbitrary – across contexts. This insensitivity led them to the use of a compensatory learner strategy of supplying an overt subject in each sentence, a rule of their interlanguage grammar which is viewed here as the product of the complex interplay between incomplete analyses of L2 input and cognitive mechanisms of explicit learning (e.g. Hawkins 2009).¹³ Language processing of this kind partly explains why long residence times and massive target language input did not improve near-native L2 performance. In sum, the findings of the L2

¹³ Note that recent neurolinguistic studies show that even L2 development without formal instruction apparently gives rise to knowledge of explicit language rules (hence explicit knowledge) since brain responses evidenced conscious processing for speakers of this kind (Batterink and Neville 2013b).

study are interpreted along the lines of a failure to reset parameters in L2 development, regardless of quality and quantity of input. Interestingly, L2 speakers with shorter residence times outperformed those with longer residence times. This result was unexpected and is interpreted as the possible impact of initial attention to massive L2 input in terms of conscious reanalysis of input and memorization. This explanation is supported by e.g. the positive effect of language training that flattens over time (e.g. Takagi 2002).

In conclusion, it is obvious that some grammatical features of L2 subjects have not been fully acquired or correctly *reassembled* in the sense of Lardiere (2007), thus impeding parameter resetting, even after long-lasting exposure to massive L2 input. Note the importance of what is taken as evidence for successful acquisition since the presence of a particular pronominal form in the interlanguage of L2 speakers does not necessarily guarantee the acquisition of its correct featural configuration (Di Domenico 2015). The persistent interlanguage deficits which were attested in the performance of the adult near-native L2 speakers point to underlying knowledge representations which are different from those of L1 speakers. I therefore argue that the interlanguage grammars of near-native speakers should be considered permanently deficient, incomparable to and fundamentally different from native grammars and thus of a non-native nature, which in turn outlines the limits of ultimate attainment in L2 development. Furthermore, since some of these interlanguage deficits are viewed as grammatical in nature (pointing to the incomplete acquisition of syntactic features of German subjects) as evidenced in deviant expletive subject use, they are a challenge for the Interface Hypothesis: expletive subjects are not subject to discourse conditions and should be invulnerable or acquirable in L2 development. Note that the fact that expletives are typically null in the L1 grammar (due to the positive setting of the Pro-drop Parameter) did not improve performance.

An outline of the content of individual chapters (now) follows this general introduction. The second chapter deals with null subject theory, introducing the most relevant theoretical contributions to the syntax of null subjects, the Null Subject Parameter and its cluster of properties, including early and contemporary proposals as well as their critical discussion. Chapter 3 introduces relevant typological information on Bulgarian (existing descriptive data) and provides arguments for the need for their empirical investigation. Subsequently, the three independent typological studies and their results are introduced and discussed. The fourth chapter introduces a considerable body of theoretical and empirical literature on language acquisition and the nature of interlanguage competence, including

some critical discussion thereof. In particular, this chapter deals with comparisons of first and second language acquisition, the nature of interlanguage grammars and important concepts in L2 theory. Chapter 5 compares Bulgarian and German subjects as part of the theoretical background necessary for the investigation of the interlanguage competence of proficient/near-native L1 Bulgarian-L2 German speakers with respect to their knowledge of foreign pronouns. An overview of German overt and null subjects as well as a cross-linguistic comparison of German and Bulgarian subjects with a focus on conflicting aspects as to their presence in the language input follows. The sixth chapter provides detailed information on the L2 study, its results and their discussion. The final chapter summarizes the main theoretical points related to the empirical investigation of Bulgarian and L1 Bulgarian-L2 German interlanguage of proficient/near-native speakers and discusses these in the light of the results obtained from the range of empirical studies carried out. These results confirm the status of Bulgarian as a pro-drop language and provide novel insights into the use of postverbal subjects in relation to the context of new information focus. The findings of the L2 study provide evidence for the nature of interlanguage grammars on the basis of data of L1 Bulgarian speakers who are proficient/near-native speakers of German as L2. Overall, the empirical findings are unique and novel, demonstrating the limits of some well-known theoretical proposals.

CHAPTER TWO

THEORETICAL BACKGROUND ON NULL SUBJECTS

Introduction

Whatever the Null Subject Parameter is, it has many subtle effects that no superficial theory will capture.
(Jaeggli and Safir 1989, 20-21)

This chapter deals with linguistic theory and null subject theory in particular. Since null subjects are traditionally viewed in relation to the Null Subject Parameter in *Generative Grammar*, the parameter itself and main concepts of syntax are introduced, including related theoretical developments. The null subject of the type *pro*, defined as the null subject in pro-drop languages, is the main focus, while additional theoretical proposals on the null and overt subject alternation or the conditions of availability and use of either null or overt subjects in these languages are also introduced and critically evaluated. Importantly, it will become apparent that differentiation between pro-drop languages and other language types is insufficient. In addition, information structure effects and the influence of the structural type of pronouns are considered. The chapter concludes with a detailed discussion of some typological aspects of subjects, revealing a number of issues, including some that have long ago been identified, as stated in the lead-in quotation of the chapter. These affect the descriptors of pro-drop languages, pointing to considerable differences between Bulgarian and other pro-drop languages on the one hand and the need for modification of the typological classifications as to the availability of null subjects across languages in general on the other hand. In particular, Müller's theoretical proposal on the impoverishment of inflectional paradigms (2006) and Holmberg's justification for the postulation of partial null subject languages (2005) as a language class, distinct from pro-drop languages are called into question. An important consequence of this discussion is the uncertainty as to the classification of Bulgarian in terms of the null subject property, which in turn makes the empirical investigation of the language necessary.