Tackling the Concept of Difficulty from Different Perspectives: The Case of German as a Foreign Language

by

Jasmin Hirschberg

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Germanic Language, Literatures and Linguistics

Department of Modern Languages and Cultural Studies
University of Alberta

Tackling the Concept of Difficulty from Different Perspectives:
The Case of German as a Foreign Language

Abstract

The German language has had the reputation of being a notoriously difficult language to learn (Twain, 1880; Sick, 2016, 2004). In addition to popular belief and, partially, research, student voices in the classroom often revolve around how “hard” or “difficult” something is with regard to learning German. In this dissertation, a multi-perspective approach to the concept of difficulty is used to operationalize what constitutes difficulty for English-speaking beginner and intermediate learners of German. The following perspectives will be in the focus: the subjective, the psycholinguistic perspective, the pedagogical, and the acquisitional perspective.

Four studies were conducted that focused on each one of these perspectives: Study 1 identified which features are perceived as difficult for beginner learners of German and for what reasons. A questionnaire was used that required students to assess the level of difficulty of 19 grammatical structures and indicate reasons for why specific structures were easy or difficult.

In Study 2, the question was addressed as to whether grammatical structures are equally easy/difficult on four measures of implicit and explicit knowledge. Implicit knowledge was assessed in the form of the oral imitation test as well as the timed grammaticality judgement test; explicit knowledge was measured with the metalinguistic knowledge test and an untimed grammaticality judgement test. Due to the time-

consuming nature of the tests and the fact that the data was collected during class time, the learner groups for Studies 1 and 2 were not the same.

Study 3 aimed to identify features of the metalanguage used in explanations of 13 grammatical structures in six beginner textbooks for German as a foreign language at the beginner level and how they may be connected with difficulty. Metalanguage was assessed by looking at the number of overall vs. distinct metalinguistic terms, the number of assumed vs. explained terms and, finally, with regard to their level of opacity or transparency (Berry, 2010).

The final study was devoted to identifying an overall sequence for the introduction of grammatical structures across both bilingual and monolingual beginner textbooks for German as a foreign Language. Textbook sequencing is closely connected to the acquisitional perspective to difficulty (Collins et al., 2009) as well as to L2 research on acquisitional sequences in non-instructed environments. In order to account for the relativity of the position of individual features and to, thereby, making point of introductions across textbooks comparable, normalization was introduced; a method that had previously not been used in research for this specific purpose.

The synthesis of the findings from these four studies leads to a comprehensive and fine-grained understanding of what constitutes difficulty in the context of beginner classes of German as a foreign language for English speakers. Based on these findings, multiple ways to decrease difficulty in the classroom are suggested. Future research should assess the practicability and effectiveness of these measures.

Preface

This thesis is an original work by Jasmin Hirschberg. The research project, of which this thesis is a part, received research ethics approval from the University of Alberta Research Ethics Board 2, Project Name “Exploring the relationship between perceived grammatical difficulty and implicit / explicit knowledge in the German as a Foreign Language Classroom”,

Study ID Pro00066728, 11/15/2016

Acknowledgements

Writing this dissertation has been one of the most challenging, yet most rewarding experiences in my life so far. Throughout the writing process, I have received a great deal of support, encouragement, and assistance.

I would first like to thank my supervisor, Dr. Leila Ranta, whose expertise was invaluable from the very first stages of the research process to the final steps of writing this dissertation. Despite the difficult circumstances resulting from long-distance supervision, you have been very accommodating, and your feedback was always extremely insightful and helpful and made sure that the quality of my work was brought to a higher level.

I would also like to thank the members of my supervisory committee, Dr. Yvonne Lam and Dr. Xavier Gutiérrez. Not only has your feedback sharpened my thinking about the studies I was working on, it has also taught me new perspectives and insights. I would like to thank you, in particular, for the promptness of your feedback in the final stages of writing this dissertation, which has allowed me to finish this project as planned.

To Dr. Mary O´Brien and Dr. Kent Lee – I would like to thank you for taking the time to serve on my committee and for the constructive, comprehensive, and valuable feedback you have provided. Your comments have helped tremendously with the finishing touches for this dissertation.

A debt of gratitude is also owed to the students, instructors and colleagues who were willing to participate in the (pilot) studies and provided feedback on the instruments, the data and coding. These processes were often very time-consuming and I am very thankful for your patience and your valuable input.

Finally, I would like to thank my family who has helped me develop the self-confidence and stamina it needs to finish a project like this. I will be forever grateful for your support.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

Chapter 1

Introduction to the Dissertation

In this chapter, the core topics of this dissertation will be introduced and reviewed within a larger practical and theoretical context: German as a foreign language (GFL) and the concept of difficulty. The chapter thereby serves to offer the theoretical and terminological foundation for the four studies that were conducted and it explains the relevance of the topic.

First, the standing of German as a foreign (and, partially, as a second) language worldwide will be reviewed. In the following section, the focus is narrowed to the role of German in North American countries with a special emphasis on the academic setting. It further establishes a connection between language learning and academic success and the relevance of the quality of instruction. After providing background information on why the German language is often described as *difficult,* a more general approach to the concept of difficulty follows.

In the final section of this introduction, it will be explained how the approaches and perspectives mentioned above apply to this thesis. The introduction concludes with a brief presentation of the four studies that were conducted.

The Case of German as a foreign language

The reasons for learning a new language are manifold: international travel, being able to communicate with people from other cultures, increasing employment opportunities and overall mobility are amongst those that first come to mind. In 2004, the UK Subject Centre for Languages compiled an impressive list of more than 700 reasons

for studying a foreign language based on comprehensive literature reviews, surveys and interviews with students from secondary schools and universities in the UK. These reasons were then categorized: some of these categories were academic skills, autonomy, careers, culture, critical thinking, identity, globalization, information acquisition, mobility, networking, study abroad, uniqueness and values. One of the main benefits of the UK study, according to the authors, is that it "will enable languages to be more effectively marketed" (Gallagher-Brett, 2004, p.2). In other words, this list is to help increase the number of students who decide to learn a foreign language.

Besides the impressive number of reasons *for* studying a foreign language, there seem to be convincing reasons *against* it as well. Language courses in general are on the decline on both secondary and post-secondary level, be it in North America (see MLA annual report 2016 or CAUTG report 2012/13) or Europe (Haeuser, 2012; Phipps, 2015). In secondary schools in the UK, for example, enrolment in German and French classes has dropped by 67% and 63% between 2002 and 2018 (Joint Council for Qualifications, 2018). For the USA, the MLA has identified a similarly considerable decline in overall enrolment since the 1960s and another, particularly high one, for most languages between the years 2013 and 2016 (e.g. 20.1 % for Italian, 20.8% for Portuguese, 11.1% for French, 13.1 % for Chinese, 7.1% for German and 9.8% for Spanish).

While German as a foreign language is no exception in regards to reports from English-speaking countries, it is clearly on the rise from a global perspective. Every five years, the German Department of Foreign Affairs publishes a report on the number of German learners worldwide, the most recent one being from 2015 (Kultusminister-konferenz KMK website, 2015). According to this report, a total of 15.4 million people

worldwide currently study German as a foreign language in institutional settings at the secondary and post-secondary level (including private institutes). This means enrolment in German courses has not been on the decline (worldwide) for the first time since the year 2000. Regions with the highest increase of German learners are the following: the Middle East, South-Western Asia, Latin America as well as Africa. The highest increase was observed for Poland, which now has more learners of German than any other country worldwide. For German, reasons for studying the language need to be assessed within a political and economic context. While the German Department of Foreign Affairs looks at German as a foreign language exclusively, the long-term motivation for studying German is probably of a more instrumental nature given the current economic stability of the country. Germany is the largest economy in the European Union and the fifth largest worldwide; as such, it is the third largest exporter worldwide with exports mainly consisting of motor vehicles and trailers (18%), machinery (14%), chemical products (9%) as well as computers and electronic products (9%).[1] As a result, Germany is both attractive to and in need of skilled workers from abroad, which makes it more likely that the underlying interest in learning German is to eventually use it as a second and not a foreign language.

The KMK report itself supports this hypothesis with a more detailed analysis of the type of German courses that students predominantly enrol in. At the post-secondary level, students clearly prefer courses that are related to their field of study, such as *German for engineers* or *Business German*. This is particularly true for universities or institutions that have bilateral agreements or exchange programs with German universities (German Department of Foreign Affairs, 2015).

The country's economic stability, as mentioned above, also finds its manifestation in the fact that the German language holds a strong international position with a relatively high number of speakers worldwide – despite the decline in enrolment in English-speaking countries. In addition to the previously mentioned 15.4 million people who currently learn German as a foreign language, approximately 96 million people speak German as their L1 or their L2 in countries where German is the or one of the official languages, such as Germany, Austria, Switzerland, Italy or Belgium (Amman, 2018, p. 170). Another 7.5 million people use German as their L1 or L2 outside of these countries, which means in places where German is not the official language. Some examples are Brazil (1.1 million), Canada (0.5 million), France (1.3 million), Russia (0.76 million), or the USA (1.4 million). The status of the German language can further be quantified by measures suggested by Graddol (1997) and others (e.g. Amman, 2018) who multiply the *per capita* gross national product of a specific country by the number of speakers of the language in question in that country. These values are then added for all countries in which the language is spoken worldwide. Thus, the economic strength of a language can be defined. Since 1975, German has continuously been amongst the five countries with the highest economic strength; up to 2005, it was even amongst the top three:

Table 1-1: Development of economic strength of the strongest languages, adapted from Amman, 2018 (p.191)

	1975		1984		2005		2009	
1	English	944	English	4.271	English	12.717	English	14.187
2	Russian	266	Japanese	1.277	Japanese	4.598	Chinese	5.379
3	German	204	German	1.090	German	3.450	Japanese	5.029
4	French	141	Russian	801	Spanish	3.204	Spanish	5.001
5	Spanish	88	Spanish	738	Chinese	2.399	German	4.257
6	Italian	78	French	669	French	2.215	French	3.109
7	Dutch	37	Chinese	448	Italian	1.207	Portuguese	1.866
8	Arabic	26	Arabic	359	Arabic	985	Arabic	1.703

In addition to the learners who *decide* to learn German as a second language (usually prior to entering the country) to pursue careers in Germany and thus further increase the language's economic strength, a considerably high number of people with very little freedom to make their own decisions now find themselves in Germany, a country they do not know with a language they do not speak. In 2016, 745.545 people came to Germany seeking refuge, mostly from Syria, Iraq and Afghanistan. Since then, the number of refugees has grown to approximately 1.3 million people (BAMF, 2020). Regardless of the regulations under which they are granted residence in Germany, that is whether or not they are allowed to work and whether their stay is temporary or permanent, they all need to be able to master the German language to become a functioning member of society.

To accommodate the unexpectedly high number of refugees in need of German classes, countless initiatives were called into life, which all pursued the aim to train people to teach German. Not only were retired German teachers asked to return and teach but

people who had never taught before went through training programs (sometimes only two to four weeks in length) so they could teach German to the refugees. In addition to these organized initiatives, volunteers (i.e. L1 speakers of German) were recruited to teach in refugee asylums or in small neighborhood groups. In other words, for accommodating these instructional needs, quantity was valued higher than the quality of instruction. However, since Germany will continue to be in need of well-trained teachers for German as a second language, be it for learners who decide to learn the language or for those who find themselves in a situation in which they cannot make their own decisions, it is of utmost importance to shift the focus from quantitative back to qualitative approaches to ensure successful and effective instruction.

Teaching German as a foreign language

While the future of German as a second language seems to be secure due to Germany's economic position, its popularity as a foreign language in North America as well as the English-speaking countries in Europe is decreasing, as previously described. The urgency of this topic is addressed in various ways, for example in the form of round tables organized by the Canadian Association of University Teachers of German dedicated to the question what can be done to secure the future of German programs at Canadian universities during the Congress of the Humanities and the Social Sciences in Ottawa in May 2015; or in the UK where measures are taken by the government to reverse this decline "or else the UK will be less competitive globally and young people less prepared for the modern world" (Fell, 2019, para.10).

Despite the economic power of the German language and its international relevance, it has not reached the status of a *lingua franca*; a role that is still held by the English language. It is therefore not surprising that foreign language enrolment is on the decline in English-speaking countries for it is difficult to justify the decision to learn a foreign language for reasons other than aesthetic or personal ones, which makes many of the 700 reasons identified by the research group in the UK invalid.

Further reasons against studying foreign languages are defined by Tinsley and Board (2013) who have identified the fear of not achieving good grades in language courses due to either its unpredictability or the huge gap between beginner and intermediate courses as a major factor for deciding against learning a language. The resulting small number of students in language classes usually leads to fewer funding opportunities and, in turn, to even lower enrolment.

In addition to reasons for students to not even start learning a new language, drop-out rates during language courses or right after achieving the minimal program requirements are of concern as well. Many studies that focus on attrition rates and the reasons for continuing or discontinuing (language) courses today go back to research from the 1950s on motivation theory and the language classroom (Meyer, 2013; Gardner & Lambert, 1959). Within the context of French at the secondary level, Gardner and Lambert (1959) investigated how attitude, aptitude and motivation are related to performance and achievement in the foreign language. Over the years, this research has developed into a sophisticated model that shows the personal, instructional and social factors of learning and how they are connected (see socio-educational model, Gardner, 2010). Parallel to the development of this model, research has been conducted on the

relationship between enrolment attrition and individual as well as institutional factors that was partly derived from Gardner and Lambert's research. The following four aspects proved to be the most significant ones in terms of enrolment attrition – motivation, academic success, instruction, and anxiety (e.g. Pratt et al., 2009; Holt, 2006; Noels et al., 1999; Saito-Abbott & Samimy, 1997; Minert, 1992; Speiller, 1988; Smythe & Gardner, 1978; Bartley, 1970; Mueller & Harris, 1966).

Wesely (2010) contextualizes these four factors within the framework of Tinto (1975) that aimed at showing the relationship between the individual and the institution.

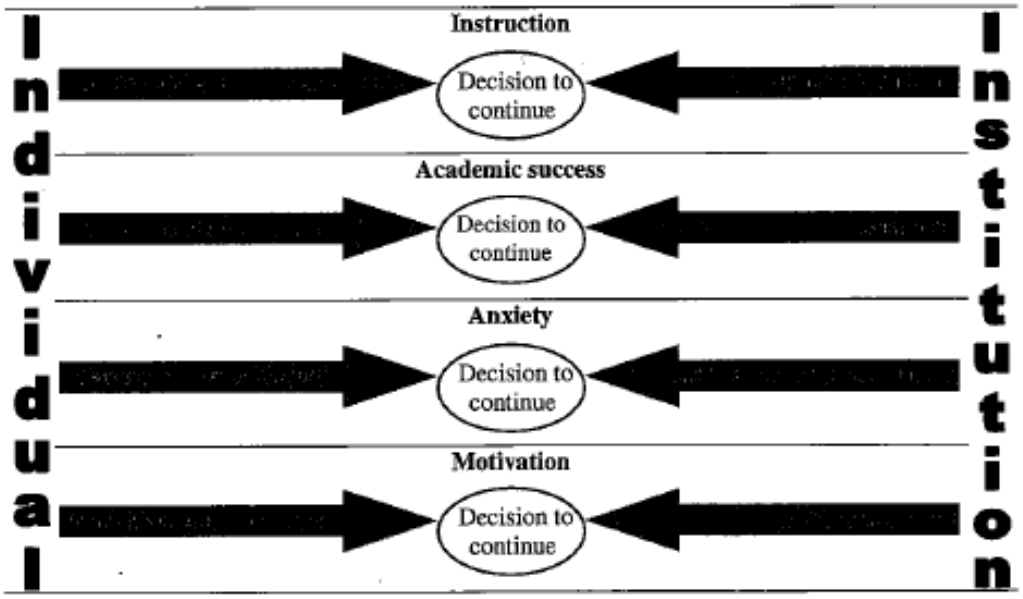

Figure 1-1: The relationship between the individual and the institution (Wesely, 2010, p. 328)

These findings make it very clear that institutions at every level have an extremely high accountability in terms of enrolment numbers. With the quality of instruction playing such a major role in a student's decision to continue or discontinue a (language) course, it should be of utmost importance to consistently ensure the highest possible quality of teaching.

The question of what constitutes *excellence* in teaching is subjective, debatable and relative. Therefore, excellence is here defined within the framework described above, that is, instruction is successful if it contributes to the students' academic success, reduces anxiety and increases motivation.

Finally, the quality of instruction largely depends on whether or not it is in line with research in the field of SLA and what we know about the acquisition of a foreign or second language. Crabbe (2003) describes research as a major source for ensuring not only learning outcomes in the form of standards, but first and foremost for "managing the quality of the learning opportunities that learners need to exploit in order to achieve the desired outcomes" (p.9). One way of determining the quality of these opportunities is through "input from research on teaching and learning" (p.32).

Despite the importance of an interaction between the two, research and practice are often perceived as two separate entities with little connection. This dissertation aims to contribute to the improvement of the quality of instruction in the field of German as a foreign language by investigating a concept that is strongly intertwined with the four aspects mentioned above, namely *difficulty*. If a course or a subject matter (i.e., a language or a linguistic phenomenon) is perceived as difficult, academic success could be jeopardized if students fail to master it and it may (therefore) further increase anxiety.

This argument also holds true for the decision to enrol in a beginner language course in the first place, as stated by Tinsley and Board (2013) and by the Joint Council for Qualifications for the UK which concludes that “the perception of languages as a difficult subject was the main reason behind a drop in the number of pupils […]” (Fell, 2019, para. 5). Depending on the students’ personality, motivation can also be affected negatively if something is perceived as too difficult.

The difficulty of the German language

Difficulty is of particular relevance with regard to German as it is, for speakers of some languages, perceived as a hard or difficult language to learn, as Mark Twain already acknowledged in his 1880 essay entitled “The Awful German Language”. In this text, he describes his own struggles as an English-speaking learner of German with constructs such as gender, separable prefix verbs, compounds and the inflection of adjectives, just to name a few. Even native speakers of German acknowledge the difficulty of their own language with the saying *Deutsche Sprache, schwere Sprache* (German language, hard language). Not only is this difficulty perceived by native German speakers, it also manifests itself in the fact that they often use cases incorrectly. For example, it is rather common among L1 German speakers to use the accusative when the dative case is needed or, even more frequently, the dative is used instead of the genitive (Sick, 2016, 2004).

A Google search on the question “Is German a difficult language to learn?” leads to 191.000.000 search results, most of which clearly answer with a “yes” to this question; for example, by describing German as “a notoriously tough nut to crack”[2] or by stating

that “I avoid languages with cases, they drive me nuts. It's just more to learn. I am surprised a language like German hasn't modernized in that regard.”[3] Other opinions are testaments of those who have stopped learning the language because “[they] did some German, but found it intolerably difficult because of [their] lack of understanding of grammar in general, let alone German grammar specifically”.[4] When asking the reverse question “Is German an easy language to learn?” over 400.000 search results come up. While the higher number of results alone may imply that public opinion on the difficulty of German is divided or even tends more towards the notion of German being an *easy* language to learn, these results are often formulated in reference to the notion of difficulty: “[…] German is not actually as hard to learn as most people think”[5], or “Many beginner level language students have the misconception that German is an almost impossible language to learn […]. However, German isn’t nearly as hard to learn as you might think.”[6]

Although these sources are by no means scientific, their effect should not be underestimated as Google is one of the most frequently used sources for *layman* research and may influence the learners’ beliefs and attitudes towards a language, which have proven to be strongly related to achievement levels in language learning (for an overview see Hsieh, 2005).

The Foreign Service Institute (FSI) of the US State Department has compiled a list of languages and their difficulty for L1 speakers of English based on the linguistic proximity between the two languages in question and on experience. The latter is based on the learning experience of FSI employees and how long it took them to learn different target languages. In that list, German is in category 2 out of 5 and therefore relatively easy and fast to learn, which seems to contradict the perceived difficulty of the language.

This study aims at contributing to the research on linguistic difficulty in the foreign language classroom by exploring which structures are *difficult* for English-speaking learners of German and how difficulty both perceived difficulty and actual measured difficulty is represented in teaching materials such as textbooks. As has been described, the reasons for learning German as a second language are manifold and will continue to grow and diversify. As a foreign language, however, it seems to be losing its appeal, which may be attributed to the common perception of German being a difficult language to learn. Therefore, research that focuses on this very concept with the objective to improve instruction and depict German as less "awful" is relevant and much needed.

What is linguistic difficulty?

The term *difficulty* is rather problematic as it is both a layman term as well as part of special registers/jargons, for example in the field of SLA. At a very general level, that is, as can be found in the Oxford's Advanced Learner's Dictionary (2000), difficulty is defined as "a problem; a thing or situation that causes problems" as well as "the state or quality of being hard to do or to understand; the effort that something involves" (p. 349). One thing that is very striking about these definitions is their vagueness, which can be attributed to one inherent feature of difficulty – its subjectivity. Lindstrom (2008) contrasts complexity and difficulty based on that very concept of subjectivity: "Difficulty […] is related to complexity but differs from it in that while complexity is a property of a system, difficulty is subjective. […] Difficulty therefore depends on the individual we take as our starting point" (p. 221). Dahl (2004) chooses the same contrast and describes complexity as "independent of use" (p. 39) while difficulty is "anchored in an agent" (p.40). Yet

another similar idea is expressed by DeKeyser (2003), who describes the above-mentioned dichotomy with the terms “objective” and “subjective” difficulty. In addition to subjectivity as mentioned above, he adds a measurable approach by describing subjective difficulty as “the ratio of the rule’s inherent linguistic complexity to the student’s ability to handle such a rule” (p. 331). Consequently, difficulty as such cannot be treated as if it exists in a vacuum but has to be analysed against different backgrounds, for example, from a contrastive perspective (i.e. language combinations), dependent on language acquisition settings (naturalistic or instructed) and processes as well as learner-inherent variables such as age, motivation or knowledge type. The following overview explores how difficulty is both defined and approached in terms of these categories in the field of SLA research.

One of the first documents that a search for literature on linguistic difficulty (especially in terms of L1 speakers of English) will find is the abovementioned report published by the Foreign Service Institute, which defines difficulty in terms of the typological distance between two languages. An accompanying article that was published by the National Security Agency (author and year of publication unknown) offers “A Comparative Analysis of Relative Difficulty” by comparing English syntax, morphology, phonology and lexicology with that of other languages and concludes that “[o]ne measure of difficulty in learning a foreign language is the degree to which the language differs from the students’ native tongue” (p.9). Rodgers (1971) adds to this dimension of difficulty, which he calls “contrastive linguistic difficulty”, the concept of “inherent linguistic difficulty”. As such, this latter field of study “focus[es] upon the physiological, maturational, and temporal factors involved” (p. 109). This description is very much in line with the one by

Dahl (2004) that focuses on the subjectivity of the concept. In 1986, an overview article on “Evaluating linguistic difficulty” was published by May, in which he uses Pienemann’s and Johnston’s (1986) order of complexity for mental processing of language as a starting point for his analysis of linguistic difficulty. While the author operationalizes difficulty against the background of different fields such as lexical, structural or conceptual difficulty, for example by considering sentence length, irregular spelling, stress or the accessibility of cultural references, a definition of the very term of difficulty is not provided. However, his list does indicate that difficulty and complexity and thereby language- as well as structure-inherent variables are not differentiated and used interchangeably.

In the approaches to difficulty by DeKeyser and Sokalski (1996) as well as Berent (1985), comprehension and production are contrasted, which means it is assumed that there may be a different level of difficulty for the learner depending on whether they have to just understand or produce the structures or features in question. While these studies focus on the learner, they do not take into account perceived difficulty, that is, a truly student-oriented approach. So far, only very few researchers have looked at difficulty from that perspective (Chavez, 2017, 2013, 2007; Shiu, 2011; Scheffler, 2009). Scheffler (2009) and Chavez (2007, 2013) used a Likert scale for students to rate a feature as “very easy” or “very difficult”, while Shiu (2011) (for Chinese learners of English) used a plethora of instruments, such as a grammatical ranking activity, interviews and stimulated recalls. However, all of these studies offer a ranking of the difficulty of certain structures from a student’s perspective, but the definition of difficulty is provided by the researcher (if at all). While Shiu does not address the issue of defining the very term in her thesis itself, she provides one on the ranking sheet for the students: “*Not at all difficult* indicates that you

have learned the structure quickly after a short explanation and practice. *Extremely difficult* indicates that you never expect to learn the structure fully, even with sufficient explanation and practice" (Shiu, 2011, p.154). Difficulty is here related to the mastery of the feature as the final measurement, while explanation and practice are perceived as facilitating factors in achieving mastery. Furthermore, this definition predominantly addresses explicit (and conscious) learning processes and outcomes. No definition or operationalization of the term is provided by Scheffler (2009) in his questionnaires; instead, he relies on the students' intuitive or layman understanding of the term. Chavez (2013, 2007) describes difficulty in terms of the likelihood to achieve an accuracy rate of 90% or higher (in speaking).

Another approach that differentiates between variables within the learner is taken by Ellis (2008, 2006), who examines difficulty against the background of implicit versus explicit knowledge. Explicit knowledge is related to "the difficulty learners have in understanding a grammatical feature" and implicit knowledge to "the difficulty […] in internalising a grammatical feature so that they are able to use it accurately in communication" (p. 88). Ellis (2009) defines explicit knowledge as conscious, declarative, accessible through controlled processing only and verbalizable, whereas implicit knowledge is intuitive, procedural, accessible through controlled processing and not verbalizable but to be observed indirectly in learners' verbal behavior (pp. 11-14). In a 2006 study, Ellis measured difficulty in terms of implicit and explicit knowledge for 17 structures of English with a participant group of 229 learners from different L1 backgrounds. His findings strongly indicate that difficulty varies depending on whether a student's implicit or explicit knowledge is tested; for example, in terms of explicit

knowledge, the third-person-singular-s as well as indefinite articles were *easy* but in terms of implicit knowledge, these structures were *difficult*.

Another line of research defines difficulty in terms of whether or not the linguistic feature is used accurately, as was briefly mentioned above (e.g. Chavez, 2013, 2007; Spade et al., 2005; Doughty & Varela, 1998). This research is summarized by Shiu (2011) as follows: “Grammar features are considered more difficult to learn if many students have difficulty using them correctly.” (p.4). However, the majority of the abovementioned studies are concerned with accuracy in one way or the other because most test batteries eventually look at the accurate use of a feature, regardless of the theoretical background which is applied. Calculating accuracy rates is how it is determined whether a feature is acquired or not. Hence, accuracy and difficulty are strongly connected.

These different theoretical backgrounds and their attempts to “[designate] a target feature as difficult” (p. 337) were categorized by Collins et al. (2009) according to the four following major perspectives: the acquisition perspective, the linguistic perspective, the pedagogical perspective and the psycholinguistic perspective. The acquisition perspective, for example morpheme order studies (Dulay & Burt, 1973) or Processability Theory (Pienemann, 1989), defines difficulty in terms of the time of acquisition, namely “whether a structure is ‘early’ or ‘late’ acquired” (Collins et al., 2009, p. 336). While a major advantage of this perspective is “that it allows actual learner behaviour to be taken into account” (p. 337), one of the drawbacks is the circularity of the argument, namely “Why is structure x difficult? Because it is acquired late. Why is structure x acquired late? Because it is difficult” (p. 338). Furthermore, difficulty can only be assessed for those

structures that have an “acquisition profile” (p. 338). Hence, the acquisition approach needs to be complemented by research that goes beyond the circularity of this approach.

The linguistic perspective focuses on the feature itself, independent from learner variables. Linguistic proximity between the L1 and L2, markedness as well as complexity are most frequently researched in this approach. As is true for any approach that does not include the learner, the findings or hypotheses that are derived from a theoretical rationale are often not in line with “expected directions” when they are tested with learners (Collins et al., 2009, p. 339). Thus, their value for the classroom and the improvement of instruction is very limited.

The more *hands-on* approach is the one that Collins et al. (2009) describe as the pedagogical perspective. As was the case for the linguistic perspective, complexity is crucial here as well. However, complexity is not considered to be a feature-inherent construct but it is defined in relationship to “the explanation for a given feature” (ibid, p. 339), that is “how amenable a rule is to explanation (by the teacher or L2 learners)”. Factors which influence this amenability are the number of exceptions to a rule, the amount of metalanguage necessary to understand and/or explain the feature or the arbitrariness of a specific feature (ibid, p. 340). German, for instance, requires a meta-language for the concept of case, that is, a learner needs to understand (and use) terminology such as *dative* and *accusative* to produce accurate output, especially in terms of adjective endings and pronouns. This metalanguage can become particularly problematic if the concept does not have an equivalent in the learner’s language.

While this approach is strongly linked with pedagogic practice, it is problematic as it usually does not account for the relationship between explicit instruction, that is

explaining a rule, and the resulting type of knowledge. Understanding a rule easily does not mean that the acquisition or the development of implicit knowledge about that rule is easy to achieve as well. For instance, if a learner understands the terminology and the concept of case in German, which may be demonstrated by elaborate explanations at a metalinguistic level, it cannot be assumed that he/she can also use this form correctly, especially under timed conditions or in spontaneous speech production.

The last approach to difficulty is described in terms of a psycholinguistic perspective within which the interplay between the learner and the input is in the focus of determining linguistic difficulty. “[W]hat makes a given language form easy or difficult is influenced by factors that facilitate or constrain learners’ access to the form in the input” (Collins et al. 2009, p. 340). Some of these factors are “unreliable and ambiguous form-function contingencies” (p. 341), redundancy of certain features or salience, that is, how perceivable a certain feature is overall.

What is shared by all of the approaches to difficulty in these perspectives is that they are lacking the answer to the core question of what constitutes difficulty; instead, they are descriptive in nature and revolve around causations such as “a structure is perceived as difficult if...”. This goes hand in hand with the fact that conclusions about the concept of difficulty is often inferred from other studies; for example, research that focuses on “[…] ultimate attainment, learnability and teachability, developmental patterns, order of acquisition, fossilization, [and] effects of instruction” (Housen & Simoens, 2016). This transfer of findings can also be seen in the overview above.

The review of the literature thus far has offered an overview of the many ways in which difficulty is approached and investigated at a general level. More in-depth

overviews with a narrower focus are provided in each of the four studies presented in this dissertation. In the following section, each study is briefly described and contextualized within the approaches to difficulty presented above.

The context of this dissertation

For the research presented in this dissertation, the following perspectives on difficulty are of particular importance: the acquisition perspective, the distinction between implicit and explicit knowledge (i.e. the psycholinguistic perspective), the pedagogical perspective, and the subjective perspective, which refers to difficulty as perceived by learners. The latter was the starting point for this project. As a teacher of German as a foreign language (or any foreign language), the concept of difficulty is of relevance on a daily basis, be it in the form of interaction with or amongst learners in the classroom. Not only do (English-speaking) learners enrol in a German class with the notion of German being a difficult language to learn, but this notion is frequently expressed in the learning process. It was the nature of these statements and the students' lacking ability to pinpoint what they mean exactly by saying that something is "hard" or "difficult" when being asked, which triggered my interest in systematically approaching the concept of difficulty. From there, the step to the dichotomy of explicit vs. implicit knowledge was taken quickly. The very inability to articulate what contributes to the learners' perception of a specific structure to be more or less difficult is a manifestation of implicit knowledge: *knowing* or *sensing* that something is difficult but not being able to verbalize it. A quick look at the literature revealed that difficulty as such does not only need to be approached with this distinction in mind (Shiu, 2011; Ellis, 2006), it also showed that this connection had not been researched for German as a foreign or second language.

As a teacher, the inclusion of textbooks and how difficulty is depicted in pedagogical materials almost happens naturally when exploring a concept that is of such relevance for both the German language itself and the learners who study it in an instructional setting. Given the fact that syllabus design as well as lesson planning heavily rely on the textbook, and thus exert a significant influence on the depiction of the language, analysing pedagogical materials was the logical final step. The result of these considerations are the following four studies. In all cases, participants were beginner (100-level) and intermediate (200-level) learners of German, which corresponds to levels A1-B1 according to the Common European Framework of References.

In the first study, the learners and their perceptions are in the focus. Based on writing samples from beginner and intermediate learners of German as a foreign language at a Canadian university, 19 structures were chosen. For these structures, learners were asked to indicate the perceived level of difficulty on a Likert scale from 1 (*not difficult*) to 6 (*extremely difficult*). In a second step, the learners had to select a representative structure for each level of difficulty (1 and 2 for *not difficult*, 3 and 4 as *moderately difficult*, and 5 and 6 as *extremely difficult*) and provide reasons for their choice. The findings indicated that learners possess a differentiated understanding of why certain structures are difficult, which is often congruent with findings in the literature.

In the second study, (most of) the structures from the first study were used to find out which were demonstrably easy/difficult for English-speaking learners of German by using measures of implicit and explicit knowledge. For the latter, a metalinguistic knowledge test and an untimed grammaticality judgement test were used to elicit data; implicit knowledge was assessed by using an oral imitation test and a timed

grammaticality judgement test. The findings from Study 2 support the claim that scores on explicit and implicit measures are different for individual structures in relation to difficulty and therefore need to be assessed individually.

The two final studies involved examining difficulty in pedagogical materials. Study 3 scrutinized the metalanguage used to explain the form and use of the same target grammatical features from the two earlier studies whereas Study 4 compared the sequence in which grammatical features are introduced in different textbooks for German as a foreign language. Metalanguage was analysed with regard to its frequency (total and in relation to distinct terms being used) and to the dichotomies *opacity* versus *transparency* and *explained* versus *assumed*. The latter indicates whether the metalinguistic terms that are used in the grammar sections are explained or part of the assumed learners' previous knowledge. In Study 4, the underlying understanding of difficulty is connected to the acquisitional perspective: easy structures are acquired early while difficult ones are acquired late. When applied to sequencing of grammatical structures in textbooks, it can be transferred that easy structures are introduced early and difficult ones later. The overall aim of this project was to identify a *general* sequence across textbooks for German as a foreign language, which was achieved by using measures of central tendency and normalization.

Finally, Chapter 6 provides a synthesis of the four empirical studies beginning with a summary of the results from each study and then a comparison of what each perspective of difficulty has to offer. While these three approaches by no means claim to reach a complete understanding of the concept of difficulty for the German as a foreign language classroom, they do contribute to answering the call for more comprehensive

approaches to the concept (DeKeyser, 2005). All three projects have implications for research and instruction, which are also briefly discussed in each paper and then, in more detail, in the concluding section.

---

1 https://tradingeconomics.com/germany
2 http://pocketcultures.com/2010/03/15/the-german-language/
3 https://www.fodors.com/community/europe/how-challenging-did-you-find-german-to-learn-950750/
4 http://donaldclarkplanb.blogspot.com/2011/02/latin-stone-dead.html
5 https://blog.busuu.com/is-german-hard-to-learn/
6 https://iwillteachyoualanguage.com/learn/german/german-tips/is-german-hard-to-learn

Chapter 2

Study 1: Exploring Learners' Perceptions of Grammatical Difficulty in the German as a Foreign Language Classroom

Chapter 2 presents a study that was conducted to examine the subjective approach to difficulty. Difficulty is operationalized as 1) grammatical accuracy in written output and 2) the perceived level of difficulty of specific grammatical structures and the reasons for these perceptions. Data was elicited in the form of writing samples and a questionnaire. Finally, it was assessed whether there was a relationship between perception and accuracy in written production.

Introduction

Learning a language is often a long and tedious journey, and some languages seem to be harder to learn than others. One aspect that influences how *hard* it is to study a specific language is the learner's L1, which is often used as a point of reference. The Foreign Service Institute (FSI) of the U.S. State Department has compiled a list of languages and their difficulty for L1 speakers of English, based on the linguistic or typological proximity between the two languages in question as well as on experience. In that list, German is in category 2 out of 5, which means it is relatively easy and fast to learn for English speakers. This assessment, however, is in conflict with less scientific but more popular and, thereby, more accessible sources. The amount of grammar that is associated with learning German still seems to be the reason for the struggles of many

learners nowadays, as the following two representative entries from a Google search on the difficulty of German show:

> "It's true that German has **four** grammatical cases, and that getting your head around the fluctuating noun and adjective endings is like mastering the tight-rope, if slightly less dangerous." (Dowd, 2018)

> "For many English speakers, German is a difficult language to pick up. Its long words, four noun case endings, and rough pronunciation gives your tongue quite the workout each time you speak." (Jumpspeak Website, 2021)

Unfortunately, the public laymen discourse on difficulty of the German language is not complemented by a very lively research-based discourse on the subject. Research that focuses on the learners' perception of difficulty is scarce in general and for German in particular. Not only does the concept of difficulty depend on the language(s) in question and their typological distance, as described above, but it is also a learner-inherent value and thereby highly subjective on an individual level. It is therefore surprising that the existing scientific discourse on difficulty almost exclusively relies on theoretical models and considerations (see Housen & Simoens, 2016; Collins et al., 2009; DeKeyser, 2005 for overviews) and dismisses learner-centred approaches to the concept for the most part. Yet, in order to adequately support the learning process and to create a learning environment in which the learner feels taken seriously on a social, cognitive and affective level, it is of utmost importance to not only focus on the objective aspects of the concept, but also on the associated attitudes and perceptions. Improving the teaching quality, which can thereby be achieved, is not only crucial for individual learning environments but also on an institutional level. Thus, the study reported here aimed to explore the often

overlooked subjective side of learning difficulty among English-speaking Canadian students in German as a foreign language classes.

A plethora of research (e.g. Smythe & Gardner, 2010; Pratt et al., 2009; Noels et al.,1999; Bartley,1970) has shown that quality of instruction is one of four main components for students to decide for or against staying enrolled in a language course. Academic success, anxiety and motivation are the other three factors. These findings make it very clear that institutions have an extremely high impact on students´ enrolment behaviour and thereby on enrolment numbers. Likewise, enrolment behaviour of students has a significant effect on individual institutions. That enrolment is a matter of concern for post-secondary institutions is revealed by enrolment statistics for North America (CAUTG report, 2015). In North America, enrolment in German courses, like many other foreign languages, has been on the decline for several years and post-secondary levels and many language programs are struggling. With the quality of instruction playing such a major role in a student's decision to continue or discontinue a (language) course, it should be among the highest priorities to consistently ensure the highest possible quality of teaching. All of the abovementioned factors are influenced negatively if a student perceives a specific subject as being too difficult. This study focuses on learners' perceptions of what is difficult about learning German grammar as a first step to build a better understanding of L2 learners' needs so as to improve pedagogical decisions about syllabus design and grammar instruction. In the following sections, the SLA literature focused on the concept of learning difficulty in general and with respect to German as a foreign language in particular is reviewed.

Literature review

Difficulty has been approached from various perspectives, resulting in numerous ways to describe and conceptualize the concept on a theoretical level. Generally speaking, difficulty can be addressed in relation to either the feature or to the learner in question (Housen & Simoens, 2016), which has been described with different labels such as “objective” versus “subjective” difficulty (DeKeyser, 2003), “linguistic versus cognitive” (Rodriguez Silva & Roehr-Brackin, 2016), or “independent of use” versus “anchored in an agent” (Dahl, 2004). DeKeyser (2005) describes the interplay between these two variables as “the ratio of the rule’s inherent linguistic complexity to the student’s ability to handle such a rule” (p. 331). The terms *rule* and *complexity* require further elaboration. The latter can be used to describe function, form or even the mapping of these two (DeKeyser, 2005). For this reason alone, researchers may refer to different things when using these terms. Complexity in this context refers to the level of a rule’s transparency, which is determined by “redundancy, optionality, and opacity” (DeKeyser, 2005, p. 8). Complexity can further entail the sheer number of rules or “the number of transformations required to arrive at the target form” (Collins et al., 2009, p.339).

The understanding of the term *rule* depends on the conceptual context it is used in: a pedagogical versus a linguistic grammar. Newby (2001) describes pedagogical grammar “as a grammar developed for learners of a foreign language” (p.1) while linguistic grammar is “theoretically” or “scientifically” describing the “set of forms and structures”, which “comprise the main focus of the textbook syllabus” (p.1). Accordingly, pedagogical rules are those that are particularly formulated for a language learner (in institutional settings). However, the interplay between a linguistic rule and its pedagogical

counterpart is not systematic since “there appears to be relatively little coherent theory underlying rule formulation” (Newby, 2001, p. 3), which often leads to “inaccurate rule formulations in […] textbooks” (Dirven, 1990). In addition, it is unclear what the relationship is between a pedagogical or a linguistic rule and the representation in the learner`s mind (DeKeyser, 2005; Robinson, 1996). Chavez (2017) even goes so far as to claim that “the basic properties of a rule are in question” (p.4) and refers to the concept of “fuzzy rules”, whose major characteristic is their approximate nature, as well as to connectionism “which takes a probabilistic view of language learners rather than the “rules and exceptions” approach” (p. 4). With these considerations in mind, Chavez defines rules and corresponding irregularities and exceptions as two categories of difficulty related to grammar-learning. In addition to these two, she further mentions 3) processing and memorization, 4) contrasts between the learner´s L1 and the target language and 5) frequency of the form in question.

In a post-secondary environment, learners have long passed the critical period and their learning-mode therefore largely “follows an explicit learning mode”, which implies that “the processing of input is understood to take place with conscious cognitive involvement and with the intention to figure out the underlying regularities” (Denhovska et al., 2015, p. 22). What is here referred to as “conscious cognitive involvement” was earlier described as noticing (Schmidt, 1990); a necessary step in order for something to be remembered. For noticing to occur (more successfully), different forms of input enhancement are desirable (Sharwood Smith, 1993). For Van Patten (1993), processing occurs when input enters the interlanguage of the learner or, as Chavez (2017) puts it, when input is introduced into memory. In contrast to memory – which requires noticing –

memorization not only requires noticing but also the conscious effort to store information for later recall. In an institutionalized learning context, memorization (i.e. rote learning) can be described as a learning strategy. Chavez (2017) points out that memorization (although it was frequently identified as crucial to the learning process in her study) is not in line with current research which "emphasizes the mental capacities that make learning possible" such as "Working Memory and its components Phonological Working Memory […] or Phonological Short Term Memory […] and Executive Working Memory" (p. 5). As will be shown later, language learners have a different understanding of what facilitates or enables the learning process.

Perceiving (grammatical) difficulty strongly depends not only on the grammar of the L1 but also on the grammar of other foreign languages that have been learned. A lack of linguistic, or typological proximity is often seen as an indicator for language learning difficulties. When comparing German to other languages, "inflectional morphology has long been seen as a central concern" (Rogers, 1987, p. 48) (e.g. for English speakers) and has therefore often been the element around which comparisons revolve (see Hawkins, 2015 for a complete contrastive typology). In fact, one of the earliest and – probably the wittiest – accounts of the German language by Mark Twain in 1880 provided a list of these frequently studied difficult features (as perceived by an English L1 speaker). With regard to grammar, he mentions the following aspects as major contributors to the difficulty of the language:

1. Case
2. Prepositions (including two-way prepositions) and use of cases
3. Sentence length and verb placement in particular
4. Separable prefix-verbs
5. Adjective endings

For this project, the contrast between English and German is of dominant relevance since most learners spoke English as their L1. As we will see, the features mentioned by Twain will also be in the focus of interest. With one of the earliest accounts of the difficulty of the German language coming from a German learner himself, it is very surprising that this approach has not been followed throughout to gain more scientific and fine-grained insight into the learner`s mind. This is even more surprising given the fact that the initially mentioned dichotomy of subjective versus objective difficulty is commonly agreed upon.

To date, very little research has actually acknowledged the subjective perspective of the very concept of difficulty by eliciting data from the learners and thereby taking a more practical and empirical approach that could contribute to more learner-centred approaches to research, as outlined above. The recently published special issue in *Studies in Second Language Acquisition* on the topic of difficulty is no exception and thereby confirms DeKeyser's (2016) claim that there is a "strong tendency to ignore" (p. 358) input from the learner. The call for acknowledging the subjectivity of the concept has, so far, largely been answered by focusing on the perceptions of the teachers (e.g. Graus & Coppen, 2016, 2015; Scheffler, 2011). While understanding teachers' perceptions of difficulty can be seen as indirect access to the students', this approach is not sufficient,

especially since the classroom experience of the teachers is largely influenced by their perspective as experts with formal education in language teaching and learning.

For German, the study by Chavez (2017) is so far the only project which elicits data from students with respect to the concept of difficulty. Chavez' objective for her study was to "[investigate] what makes learning German grammar difficult from the perspective of learners" (p.6). With four research questions, she finds out (1) which difficult items were accompanied by an explanation and (2) how was difficulty explained; 3) if the proficiency level of the learners and 4) the item itself have an influence on the nature of the explanation. A total of 96 students participated in the study who covered a range of 1st-year to 4th-year. The study was conducted at a "large Midwestern research university [in the U.S.] in a German department" (p.6). In a quantitative and qualitative questionnaire, most explanations were provided by 2nd- and 3rd-year learners with 76.93% and 88% respectively. Chavez does not provide any information on the language background of the students, which means it is not known if German is their L2 or if they know other languages as well. The fact that she only mentions differences between English and German might imply that most of the participants learned German as their L2. Out of the 96 students, 63 provided 82 explanations for why they perceive a certain feature as difficult. From these explanations, 8 response categories emerged, namely *memorization*, *complexity*, *opacity, number and lack of rules*, *slow processing*, *lack of experience* and *difference from English*. With 27% and 25% of responses, *memorization* and *complexity of rules* account for the largest number of reasons provided for grammatical difficulty, followed by *speed of processing* (13%). All other responses were below 10% (*opacity of rules, difference from English, lack of experience, number of rules*) and 5% (*lack of rules*).

One of her major findings was the fact that the terminology that the students used was identical to terms that are used in research. However, “a shared metalanguage between students and teachers/researchers not only covers up stark divisions in outlook, it also perpetuates it” (p.9). She concludes her findings as follows: “many learners […] tend toward a mechanistic view of language learning with an emphasis on memorization and a preference for formula-like rules that are invariable and connote an exact one function – one form relationship” (p.17) – in short, she describes the learners´ perception of their own minds while learning a foreign language as a “machine” with language being the “program” (p. 17). Chavez formulates very few pedagogical implications; instead, she assesses whether what learners perceive is congruent with theoretical considerations and empirical findings of the features in question, which – not surprisingly – is not the case. Therefore, she suggests that “1) students need to better understand language learning processes as well as their connection to language teaching practices [….] and that 2) […] learners need to realize that their personal theories of language learning are just that – and that research-influenced teaching methods may not accommodate them” (p.19). Her overall solution is “to share a vision”, that is, making sure learners are aware of the processes involved in language learning, partially manifest in the form of terminology and metalanguage, and how they influence pedagogy.

As in Chavez' study, the present study addressed the question of German students' perceptions of learning difficulty, partially replicating her research design in terms of the type of institutional setting and participants as well as the approach to data collection and coding. The following research questions will be addressed in this study:

1. What structures are difficult or easy to use accurately in written production of English-speaking beginner and intermediate-level learners of German as a foreign language?
2. Which structures are perceived as easy/difficult by these learners?
3. Is there a relationship between learners' perception and their production?
4. What are the reasons for perceiving a structure as easy or difficult?

Method

Participants and research setting

The study was conducted with beginner (100-level) and intermediate (200-level) university learners of German as a foreign language in a major research university in Western Canada. Enrollment in 3rd and 4th-year courses was considered as too low to allow for representative results for these learner groups.

For the beginner learners, the data was collected at the end of their second term of German, for the intermediate learners at the end of their fourth term. Data from 82 beginner and 30 intermediate learners whose L1 was English was used for the study; L1 speakers of the following languages participated but were excluded from the data analysis to ensure validity and comparability of the findings: Chinese (8), Korean (4) as well as one speaker each for Vietnamese, Hindi, Punjabi, Persian, French, Spanish, Polish and Lithuanian. Four speakers indicated that they had grown up bilingually English/French (3) and English/Spanish (1). For 37 speakers, German was the L2, 42 learned it as their L3, 26 as their L4 and 2 as their L5 and one as an L6. Given the fact that the study was conducted in Canada, this high number of multilingual individuals is not surprising since linguistic diversity at the post-secondary level is the norm. The country has official

bilingual status and every year, 50,000 officially bilingual students graduate from Canadian high schools (Saul, 2013). At the university where the present study was conducted, approximately 8,000 out of 39,000 students were on an international study permit. Not only does this diversity imply dealing with more experienced language learners at the post-secondary level, it also increases the chances of facing learners that are more motivated and possess a higher level of aptitude.

When using the term *beginner learner* in this study, it has to be mentioned that only very few participants started the beginner class with no prior language knowledge at all. The vast majority of learners (84 out of 112) had been exposed to German in one way or the other, mostly through family members, friends or partners who are L1 speakers of the language. While this does not mean that the label *beginner* does not apply to them proficiency wise, having had prior exposure to the language may have an influence on the learners' perceptions.

Based on my personal experience of having taught these and similar classes at this very institution, it is likely that the instructional style of these courses followed the principles of communicative language teaching (Brandl, 2008) on all levels: tasks are at the centre of instruction as well as the promotion of "learning-by-doing" approaches. The choice of textbooks is partly based on how meaningful and comprehensible the input is. As part of the instructional design, cooperative forms of learning are essential. While explicit instruction is used minimally, it does have its room in the daily instructional routine. At the lower level, grammar is usually embedded in communicative contexts and mainly relies on chunk-learning, while the 200-level learners are often required to review structures or study new features on their own with the help of the textbook before they

are briefly covered in class. The curriculum is very dense, that is, one textbook unit needs to be covered in one week of classes or four sessions, which equals up to 45 linguistic topics in one term (at the beginner`s level). This density is often in conflict with communicative approaches since forms of assessment usually require students to actively use these structures, that is, master them and not just offer opportunities for exposure. However, due to curricular requirements that need to be met by the individual instructors such density is necessary. In order to reduce input load during class time, written practice exercises such as gap-filling are usually assigned as homework and can be checked by the students themselves with the answer key in the textbook.

Instruments

The writing task

The writing task was designed in order to have learners produce written output that could be used to compute accuracy rates for the use of individual structures. Accuracy rates are considered to be an indicator of difficulty, that is, the higher the accuracy rate, the less difficult the structure in question and vice versa.

The writing samples were elicited from the same learners who, later in the term, filled in the Perceptions of Difficulty questionnaire. The 100-level learners were asked to write a letter to a fictional exchange student; they had to introduce themselves and their daily routines. Instructions were provided in English:

> *"You have applied to an exchange program between Edmonton and Berlin. Write a letter to your exchange partner who comes to visit Edmonton in April 2016. Introduce yourself and talk about the following topics."*

Some of the topics were: field of study, hobbies/leisure time activities, talent and duties, family, Edmonton. At first, it was planned to have the learners write the texts at home; however, in order to ensure that no resources such as dictionaries, translators or help from other people would lead to invalid data, the writing was done during class time. While no minimum number of words were requested, the task was formulated in a way that it was clear that students were asked to address all of the topics mentioned in the instructions, which would ensure texts lengths that were appropriate for the analysis. For this reason, learners had 45 minutes for the task to be completed.

The 200-level learners watched the silent movie “Good Times” (Chaplin et al., 1936) and were asked to rewrite the plot during class time for the same reasons provided above. The instructions asked them to provide as much detail as possible in their re-written version of the movie and to focus on both the characters and how they are depicted as well as the plot. After watching the movie once, they had 35 minutes to write their texts. Again, no minimum number of words was requested.

These tasks were designed to ensure that, on the one hand, the learners´ proficiency level was met. The assessment of the proficiency level was based on the course they were enrolled it at the time of data collection, namely the 100- or 200-level. The course instructors were asked prior to the data collection if there are any students enrolled in the courses that can be considered outliers in terms of their performance. One student in the 200-level course had spent more than six months in German-speaking countries; however, the instructor did not notice differences in written performance compared to other students. Therefore, this student was not excluded from the data collection and analysis.

On the other hand, these tasks were formulated to create a setting that elicits the use of a variety of grammatical structures. It thereby did not only serve to collect valid data, it also served as an opportunity for the students to review / practice what they had covered in class up to that point, which also offered further justification for the data to be collected during class time.

The writing samples were then used to calculate accuracy rates for each structure, which were later correlated with difficulty ratings. For calculating the accuracy rates, the following steps were taken for each group:

1. All structures that had been taught up to the point of the data collection were identified by looking at the chapters that had been covered in the textbooks; the structures were then confirmed with the program coordinator. An overall number of 41 grammatical structures was identified for both groups, that is, these were structures that had been taught to both groups at the time of the data collection. The samples that yielded structures that had only been part of the 200-level curriculum were not part of the analysis. For the 100-level learners, the samples only occasionally yielded structures from the 200-level curriculum and were excluded as well.
2. In all texts, the occurrence of each of these structures (or lack thereof where they should have been used) was identified. Structures that appeared – on average – less then five times in a text for both groups were excluded from the analysis for a lack of representativeness. After this step, a total of 23 structures remained with an average occurrence that ranges between 5 (case with two-way preposition) and 12 (subject-verb agreement). The features were grouped

according to their linguistic domain, namely morphology, syntax, semantics, etc. Although domains are not considered to be discrete, that is, structures do not strictly belong to one domain only, the categorization here was made in order to be able to focus on what is relevant about the individual structure regarding the way it is taught and, as a consequence, for how to code it (see below).

3. For each feature, a rigorous right or wrong coding was applied (with regards to the linguistic domain the feature had previously been assigned to). For example, the choice of the correct auxiliary (*haben* vs *sein*) was considered to be a semantic feature since the decision is based on what the main verb expresses, namely movement, change of place or state (*sein*) vs. verbs that do not express movement or a change of state (*haben*). When a participant chose the correct auxiliary but made a mistake that was not related to semantics, that is, spelling or conjugation, it was still coded as correct in these instances.
4. Spelling mistakes were ignored unless they affected the structure. For example, when coding all instances of adjective endings, spelling mistakes that occurred in the stem of the adjective were ignored and the adjective ending was still marked as correct. When a student wrote “ein **sh**öner Tag” (a beautiful day) instead of “ein **sch**öner Tag”, this instance was marked as correct as the missing c is not part of the adjective’s ending.
5. In a final step, accuracy rates were calculated; accuracy rate was defined as the percentage of correctly classified instances amongst all occurring instances. Differences between the mean accuracy rates for the 100- and 200-

level learners were calculated with a two-tailed independent t-test with a significance level of 0.05.

Based on the writing samples, 19 structures were chosen for being used in the perception questionnaire. The decision to exclude 4 of the 23 identified structures (as explained above) was based on the fact whether the structures that were used here were also part of Chavez' (2016) study. Furthermore, the 19 remaining structures were rather diverse since they covered the linguistic domains syntax, semantics and morphology; the number of items for the questionnaire was not supposed to be reduced further for reasons of representativeness. Table 2-1 shows all structures that were used, including examples.

Table 2-1: Structures chosen for analysis, presented according to linguistic domain including example

Verb morphology	**Noun (phrase) morphology**	**Syntax**	**Semantics**	**Adjective Endings**
Participles in the past tense *Ich bin **ge**laufe**n**.* *I ran.* *Ich habe **ge**kauf**t**.* *I walked.*	Accusative / dative as direct and indirect objects *Ich gebe **meiner Schwester ein Buch.*** *I give my sister (dative) a book (accusative).*	Verb placement (in dependent +independent clauses) *Ich **frühstücke** (2nd position) immer um 0730 Uhr.* *(I always have breakfast at 0730 a.m.)* *Ich frühstücke, weil ich Hunger **habe** (final position)* *I am having breakfast because I am hungy.*	Choice of auxiliary in the past tense *Ich **habe** gelesen.* *I read.* *I **bin** geflogen.* *I flew.*	*=number/gender/case agreement* *Ich habe ein grün**es** Auto, eine grün**e** Jacke und ein**en** grünen Hut.* *I have a green car, a green jacket and a green hat.*
Subject-verb agreement ***Er** geh**t** einkaufen.* *He goes shopping.* ***Wir** geh**en** einkaufen.* *We go shopping.*	Possessive adjective agreement (gender, number, case agreement) *Sein**e** Jacke/se**in** Auto ist blau.* *His jacket/his care is blue.*	Modal verb placement *Ich **muss** (2nd position) heute **aufräumen** (final position)**.*** *I have to clean today.*	Choice of preposition *Die Katze isst **auf** dem Sofa.* *The cat is eating on the couch.* *Das Messer ist **in** der Schublade.* *The knife is in the drawer.*	
	Case with two-way prepositions *Ich gehe **in die** Küche (accusative for direction)* *I am going to the kitchen.*	Adverb placement (time and place) *Mein Deutschkurs ist **jeden Abend** (time) **in Raum 7** (place).*	Choice of modal verb *Ich **muss/darf/soll** ins Kino gehen.* *I must/have to/shall go to the movies.*	

	*Ich bin **in der** Küche (dative for location)* *I am in the kitchen.*	*My German course runs every night in room 7.*		
	<u>Determiner agreement</u> ***Der** Deutschkurs (masculin)* *The German class* ***Die** Klasse (feminine)* *The course* ***Das** Kursbuch (neuter)* *The course book*	<u>Separable-prefix verbs</u> Ich **nehme** an der Konferenz **teil**. (verb: teilnehmen) I participate in the conference.	<u>Choice of possessive adjectives</u> ***Mein/dein/sein/ihr** Haar ist sehr lang.* *My/your/his/her hair is very long.*	
			<u>Definite / indefinite article</u> ***Der** Mann ist groß.* *The man is tall.* ***Ein** Mann ist groß.* *A man is tall.*	
			<u>Negation</u> *Ich habe **kein** Auto.* *I don´t have car.* *I möchte **nicht** ins Bett gehen.* *I don`t want to go to bed.*	
			<u>Two-way prepositions</u> *Die Schuhe liegen **unter/auf/neben** dem Sofa.* *The shoes are under/on/next to the couch.*	

Perceptions of difficulty questionnaire

The perceptions of difficulty questionnaire was designed for this study based on the structures that were identified in the writing samples as described above. It elicited both quantitative and qualitative data for analysis. The first section asked students to rate the 19 features in question on a scale from 1 (*not difficult at all*) to 6 (*extremely difficult*) (see full list in Table 2-1):

> *"The following is a list of German grammatical features. Please indicate on a scale from 1-6, whether a particular grammatical feature has been more or less difficult by circling only* ***one number****."*

A six-point scale was used in order to avoid a middle point or neutral answer. While verbal descriptors for each value were not included in the instructions of the first part of the questionnaire, they were provided for all six values in the second part (which had to be read before working on the first section). Oral instructions further drew the participants' attention to the fact that values 3 and 4 indicate *moderate difficulty*. The middle values 3 and 4 are therefore not considered to be neutral or non-committal answers but a representation of moderate difficulty. This assumption was supported by the fact that the data did not provide evidence that learners misinterpreted the scale from 1-6 in the first part of the questionnaire, that is, answers were distributed quite evenly across categories (exceptions see the result section) and no accumulation could be observed for the middle values.

Since the 100-level learner groups (in their second term) and the 200-level learners may have received German instruction elsewhere prior to enrolling in the courses that participated in this study, the questionnaire offered the option to check that a feature *had*

not been taught yet. As was true for the comparison of average accuracy rates for 100- and 200-level learners, a two-tailed independent t-test with a significance level of 0.05 was used to compare the average levels of perceived difficulty.

In the second section of this questionnaire, participants were asked to explain why a specific feature was either *not difficult at all* (1 or 2 on the scale), *moderately difficult* (3 or 4 on the scale) or *extremely difficult* (5 or 6). The three different levels were added to the analysis in order to allow for a more fine-grained analysis and in order to take into account the reasons for a structure to be perceived as *easy* or *difficult*.

For each category (i.e. *not difficult at all*, *moderately difficult* or *extremely difficult*) the students had to choose one feature and provide reasons for their choice:

> Please indicate which of the above mentioned feature was **not difficult at all (1 or 2 on the scale)** and explain why. Please explain in detail.
>
> I ranked feature __XY__ (only one) as **not difficult at all** because:

In this study, participants had to choose three specific structures and explain why or why not they were difficult. Chavez (2017) asked the learners to provide reasons for all structures that she included in her study. While requesting the participants to provide explanations for all 19 structures would have resulted in more data overall, focusing on only three structures was considered to increase the chances of eliciting more comprehensive and thorough responses. Reflecting upon reasons for 19 structures to be difficult would have been very time-consuming and too demanding. However, it had to be taken into account that focusing on only three structures would limit possible answers,

which led to data that provides insight into general reasons for difficulty and easiness instead of data that allows for information on feature-specific reasons.

The question format was open and the terminology that was used was the same as the one provided in the textbooks of these courses. Examples for each structure ensured that students knew what the term in question meant. A definition of the term *difficulty* was not provided in the questionnaire in order to avoid bias in the participants, which would have compromised the research question.

Analysis

For the analysis of the qualitative data, all students´ explanations were catalogued and then coded based on the information units (Lincoln & Guba, 1985) they contained. A response (i.e., explanation), could contain more than one information unit, which is why the latter were used for cataloguing purposes. The following example illustrates the difference between responses and information units:

Adjective endings:

> "It's just so much stuff to memorize, which is hard because I don't even understand the rules. They were not properly introduced in class either."

While this sentence could count as one response, it contains three different units of information. After coding the data for these units in a first step, they were later categorized according to their informational content or themes. In the example from above, these themes are *memorization* ("so much stuff to memorize"), *rules* ("I just don´t understand the rules"), and *instruction* ("not properly introduced in class"). Only data from

this study was used to identify these information units, that is, all themes emerged from the responses provided by the students in this study and not from Chavez' study. For the identification of the themes, it was considered as crucial to strongly rely on the students' responses instead of *interpreting* the data. Therefore, themes were often identified and named based on the wording in the students' responses. For example, "memorization" or "memorize" was often explicitly mentioned in the responses, which was then used as a theme (i.e. "memorization"). Examples as well as brief operationalization for each identified theme can be found in Table 2-6.

In order to achieve the highest amount of specificity, information units were catalogued instead of responses in all cases. For a comparative analysis between the two groups, the same structures were used for both the beginner and the intermediate learners. To ensure consistency, only data from participants was used where there was an overlap between the rating of the features in the first part and the second part of the questionnaire. For example, if a participant rated agreement with a 1 (= not difficult) in the first section but in the second section used agreement as an example for a moderately difficult structure (which corresponds to a value of 3 or 4), this data was not included. Also, if a student did not rate any structure as *extremely difficult* (i.e. a value of 5 or 6 on the Likert scale) in the first part of the questionnaire but then did pick one of the structures that were assessed to be *not difficult* or *moderately difficult* as an example for *extreme difficulty* in the second part of the questionnaire, this data was excluded from the analysis as well. Likewise, if a structure was rated as *extremely difficult* in the first part but in the second part no example and explanation for an extremely difficult structure was provided, the data was not taken into account either. Overall, instances that showed such

inconsistency occurred only 7 times, which further supports the assumption that the scale from 1-6 was not misinterpreted.

Results

Grammatical accuracy in written production

Table 2-2: Mean accuracy rates, *t*-values and *p*-values, 100- and 200-level learners, sorted alphabetically

Feature	**Mean 100-level learners**	**Mean 200-level learners**	***t*-value**	***p*-value**
Accusative	0.54	0.55	-0.19	.842
Adjective endings	0.35	0.52	-0.61	.543
Adverb placement	0.70	0.75	-0.10	.919
Agreement	0.73	0.62	-0.28	.780
Auxiliary choice in the past tense	**0.49**	**0.58**	**-2.51**	**.014**
Case with two-way prepositions	0.33	0.43	0.02	.984
Dative	**0.11**	**0.29**	**3.26**	**.002**
Definite / Indefinite article	0.82	0.86	1.39	.167
Determiner agreement	**0.53**	**0.73**	**-4.29**	**.000**
Modal verb choice	0.87	0.90	1.42	.663
Modal verb placement	0.80	0.81	-1.34	.183
Negation	**0.75**	**0.96**	**-2.70**	**.008**
Participle choice in the past tense	0.43	0.58	0.69	.487
Possessive adjective agreement	0.42	0.51	1.58	0.327
Possessive adjective choice	0.86	0.88	0.29	.382
Preposition, Choice of	0.47	0.47	0.01	.989
Separable-prefix verbs	0.77	0.80	-0.47	.638
Two-way prepositions	0.75	0.63	0.39	.693
Verb placement	0.59	0.77	-1.48	.141

The values in Table 2-2 show the ratio between the correctly classified instances and all occurring instances. A value of 0.54 means that – on average – learners were 54% accurate when using the accusative in the output that was elicited by the writing task described above.

Overall, the 200-level learners reached a higher mean accuracy rate for each feature than the 100-level learners. The only two exceptions are subject-verb agreement and two-way prepositions: the 100 level-learners achieved an accuracy average of 73% and the 200-level learners 61% for agreement and 75% versus 63% for two-way prepositions. However, a statistically significant difference between the accuracy mean values for the 100- and the 200-level learner groups existed for only four structures: auxiliary choice, dative, determiner agreement, and negation (in bold).

All of the more difficult features are related to noun (phrase) morphology and adjective endings. In comparison, the fields of verb morphology, semantics/conceptual and word classes have relatively high accuracy rates as depicted by the following grouped representation of the data:

Table 2-3: Grouped mean accuracy rates, 100- and 200-level learners

Feature Group	100-level learners	200-level learners
Adjective Endings	0.35	0.53
Noun Morphology	0.38	0.50
Verb Morphology	0.58	0.59
Syntax	0.70	0.78
Semantics	0.72	0.76

Grammatical structures perceived as relatively easier/more difficult

Table 2-4: Means, *t*-values and *p*-values for perceived difficulty, 100- and 200-level learners, sorted alphabetically; values range from 1 and 2 (*not difficult*), over 3 and 4 (*moderately difficult*) to 5 and 6 (*extremely difficult*)

Feature	Mean 100- level learners	Mean 200-level learners	*t*-value	*p*-value
Accusative	2.34	2.40	-0.26	.796
Adjective endings	3.66	4.23	-1.87	.064
Adverb placement	**2.23**	**1.60**	**-2.96**	**0.003**
Agreement	1.61	1.63	-0.11	.909
Auxiliary choice in the past tense	**2.37**	**1.63**	**3.37**	**.001**
Case with two-way prepositions	3.00	3.27	-0.94	.349
Dative	2.85	2.93	0.49	.618
Definite / Indefinite article	1.70	1.83	-0.54	.587
Determiner agreement	2.70	3.33	1.89	.608
Modal verb choice	2.18	2.10	-0.33	.373
Modal verb placement	2.26	1.93	1.24	.218
Negation	2.26	2.00	-0.37	.715
Participle choice in the past tense	2.99	2.67	-1.19	.237
Possessive adjective agreement	2.49	2.83	1.33	.183
Possessive adjective choice	2.56	2.53	0.18	.847
Preposition, Choice of	2.67	2.42	-0.16	.871
Separable prefix-verbs	2.15	1.97	-0.82	.417
Two-way prepositions	**1.98**	**2.89**	**3.41**	**.000**
Verb placement	**2.74**	**1.80**	**3.69**	**.000**

With a mean value of 3.66 (100-level) and 4.23 (200-level), article endings are by far the structures that were perceived as most difficult, followed by determiner agreement, case with two-way preposition, dative and accusative cases. Verb placement and adverb placement, two components that were categorized as syntactic were, however, not perceived as particularly difficult, especially by the 200-level learners, with a mean value of 1.80 for verb placement and 1.60 for adverbial order (time before place). The 100-level learners perceived it as more difficult (2.74 for placement and 2.23 for time before place). Surprisingly, most structures were perceived as *not difficult* or *moderately difficult*. None of the structures were perceived as *extremely difficult* according to the group mean. The range for individual answers was also fairly narrow, that is, there was little variation in the individual answers.

When the 100- and 200-level learners are compared statistically, it is worth noting that, while the 100-level learners show a slightly higher mean difficulty rating, for most structures this difference is not significant. A significant difference only exists for adverb placement, auxiliary choice, two-way prepositions and verb placement (in bold). Comparisons were made by using a two-tailed independent t-test, with a significance level of 0.05.

Relationship between accuracy rate and perceived level of difficulty

In order to determine the relationship between accuracy rate and perceived difficulty, Spearman's rank-order correlation coefficients (r_s) and significance (p) were calculated for each feature. Correlation is significant at the 0.05-level (2-tailed).

Table 2-5: Spearman Correlation Coefficient and significance for association between perceived level of difficulty and accuracy rates, 100-level and 200-level learners, sorted alphabetically

Perception		Accuracy	
		100 level	**200 level**
Accusative	Spearman Correlation Sig. p (2-tailed) n	-.007 0.949 76	.007 .970 30
Adjective endings	Spearman Correlation Sig. p (2-tailed) n	-0.006 0.955 80	.014 .942 29
Adverb placement	Spearman Correlation Sig. p (2-tailed) n	-.167 .190 63	-.180 .519 22
Agreement	Spearman Correlation Sig. p (2-tailed) n	.017 .878 82	-.521 **.004** 28
Auxiliary choice in the past tense	Spearman Correlation Sig. p (2-tailed) n	.039 .759 65	.184 .339 30
Case with two-way prepositions	Spearman Correlation Sig. p (2-tailed) n	.100 .377 80	-.009 .959 30
Dative	Spearman Correlation Significance p (2-tailed) n	.008 .948 76	.181 .339 30
Definitive / Indefinite articles	Spearman Correlation Sig. p (2-tailed) n	.062 0.589 80	.059 .786 29
Determiner agreement	Spearman Correlation Sig. p (2-tailed) n	.145 .206 79	.196 .299 30
Modal verb choice	Spearman Correlation Sig. p (2-tailed) n	.205 .149 56	.519 **.013** 22
Modal verb placement	Spearman Correlation Sig. p (2-tailed) n	-.029 .831 56	.136 .546 22
Negation	Spearman Correlation Sig. p (2-tailed) n	.303 **.013** 66	-.388 **.045** 27

Participle choice in the past tense	Spearman Correlation Sig. p (2-tailed) n	.200 .107 65	-.032 .865 30
Possessive adjective choice	Spearman Correlation Sig. p (2-tailed) n	.004 .975 78	.067 .724 27
Possessive adjective agreement	Spearman Correlation Sig. p (2-tailed) n	-.060 .599 78	.096 .615 27
Prepositions, Choice of	Spearman Correlation Sig. p (2-tailed) n	.008 .945 80	.048 .799 30
Separable prefix-verbs	Spearman Correlation Sig. p (2-tailed) n	-.009 .946 64	.012 .952 27
Two-way prepositions	Spearman Correlation Sig. p (2-tailed) n	.200 .075 80	.083 .663 30
Verb placement	Spearman Correlation Sig. p (2-tailed) n	-.301 **.005** 82	.038 .843 30

With the exception of the values for negation and verb placement, no significance could be identified for the association between accuracy rates and learner perception for the 100-level learners, which means that the Null Hypothesis was confirmed. For the 200-level learners, the association was significant for agreement, modal verb choice and negation, which means H_1 was confirmed. In all cases, the correlation is positive: the higher the accuracy rate, the higher the perceived level of difficulty.

Learners' reasons for their ratings of learning difficulty

For the 100-level learners, 363 information units were identified and 140 for the 200-level learners for a total of 504. With regard to the three categories, there was a similar distribution of the three difficulty levels, meaning that they wrote similar amounts for each category: 176 information units for not *difficult*, 171 for *moderately difficult* and 157 for *extremely difficult*. Per feature, the number of themes that were mentioned varied between six and nine. The themes were distributed as follows: for the majority of the information units, the two or three most frequently mentioned ones accounted for the majority of the information units, making it possible to clearly identify the major contributors for either perceived easiness or difficulty. Overall, the explanations were fairly distinct, which made it possible to identify clear reasons for general easiness and difficulty, but they also reflect the nature of the grammatical feature in question. For example, *memorization* was one of the most frequently identified themes, which was often explained by the students by listing what it is that needs to be known and memorized; for adjective endings, for example, the type of article, gender and number were mentioned in addition to the corresponding endings themselves. In explanations for separable prefix-verbs, learners mentioned that one either needs to memorize which verbs belong to that category or one needs to know how to find out if it does. In addition, the explanations further included that the prefix moves to the final position of the sentence, which was often "made very clear" in class.

Table 2-6 displays the themes that emerged from the qualitative analysis in this study based on the responses provided by the participants. A comparison of the findings from this study and Chavez' study follows in the discussion section.

Table 2-6: Reasons for perceived difficulty, all learners combined

Theme Present Study	Example *not difficult*	Example *difficult*	Operationalization
Memorization and Learning Strategies	"All it takes is memorizing; that's easy"	"It's just too many things to memorize and I don't know how."	Themes were mostly explicitly mentioned in responses (e.g. "memorizing", "strategies to memorize)
Processing Time / Accessibility	"I don't even have to think about. I know it right away."	"It takes too much time to come up with the right ending when I want to say something."	Theme emerged from reference to the dimension of *time* to access a form (and use it)
Cross-Linguistic Transfer	"I'm used to referring to objects by gender from French so it's easy for me to do the same in German"	"English doesn't attach any endings to an adjective, so it's a really foreign concept to me."	Explicit comparison of German with other languages
Instruction	"It has been hammered into our heads so many times that it's a piece of cake for me now."	"We don't practice separable prefix- verbs enough in class."	Responses referring to teacher actions (mostly explanations and amount/quality of exercises) in the classroom
Cognition	"I just got it right away."	"I just don't know how to do it."	Responses referring to *understanding* the nature and use of a specific feature
Rule Consistency	"Once you get it, it's always the same."	"There's not even a point to learning it because there are too many exceptions."	Responses referring to regularity/ exceptions to specific rules
Opacity of Rules	"The difference between definite and indefinite articles is just very clear; there is never a shadow of a doubt"	"Sometimes it's not really clear what counts as a motion verb and what doesn't."	Theme emerged from explicit reference to "clarity" or lack thereof
Complexity of Rules	"It's just one question you need to answer: is it a direction or a location? That's it."	"There are so many things you need to take into account when you want to find the right ending that it's almost impossible to do."	Responses referring to the number of aspects that need to be considered in order to arrive at the correct form
Lack of Rules	-	"Gender is just random, which makes it really hard."	Theme was explicitly mentioned in responses: "random", "there are no rules"

As the example comments show, the categories are applicable to both easy and difficult structures. *Memorization*, for example, is perceived as easy as long as it does not become too much to memorize or if the students lack strategies for doing so, which is usually attributed to a lack of practice and the provision of memorization strategies in class: “I wish they taught us some strategies to memorize like mnemonics, for example.” On the spectrum of *processing time*, difficulty is predominantly described as either “not intuitive” and not requiring “any thinking” or in terms of the time it takes to access the knowledge in question when asked to produce output. The latter is identical with Chavez’ category *slow processing*. *Easy* features are those that are “intuitive” or “second nature” compared to those where “intuition doesn’t work” or where things “run against what feels natural”. In many cases, the term was operationalized by the students themselves by describing something to “sound right” or “feel right” and “it just seems to make intuitive sense to me”.

Explanations for *cross-linguistic transfer* included English as well as Spanish and French. Other languages were not mentioned though there was a high variety of other languages learned. For the *easy* end of the spectrum, English was described as a source for positive transfer and thereby a support for the learning process. On the *difficult* end of the spectrum, lack of parallels between English and German was mentioned as a source for problems. English and German were contrasted on three different levels: conceptually (“There is just no such thing as gender in English really”), in relation to memorization effort (“It is so easy to remember because it’s almost identical in English.”), and rule consistency (“In English, there is only one rule to follow but here, there are too many exceptions that it gets really unpredictable”). With one exception, Spanish and French were exclusively

mentioned as facilitators of the learning process and thereby at the *easy* end of the spectrum ("Gender is easy for me because it is the same as in French"). As was the case for English, comparisons were made in terms of memorization effort, rule consistency, and conceptually.

Instruction was another relevant contributor to difficulty. The aspects that were brought up here were explicitness of instruction, nature and frequency of practice, and time of introduction of a new feature. A lack of practice and explicit instruction make students perceive a feature as more difficult while frequent practice and explicit instruction (including an explanation of the terminology and explicit instruction in English, not German) contribute to the perception of a grammatical item as *easy*: "I wish they would teach us some grammar in English before we're required to understand German.", "I hate it that we're never told what is going on. I don't like guessing games – just tell us what we need to know and be done with it already" or "I have the feeling we never practice it in class. We have tons of homework but I really wish we'd actually be using the past tense in class." The nature of practice that was described as most effective were drills and other repetitive forms of exercises: "We have done it so many times in class that it now feels automatic." or "We did it over and over and over again and now I dream about it, which is good." Features that were introduced just recently were, therefore, perceived as more difficult than those that were taught in the beginning, as this comment shows: "It was one of the first things we ever learned, so it's really easy now".

Cognition, on both ends of the spectrum, includes comments that are not further explained but make it clear that understanding the feature in question either exceeds the cognitive abilities of the learner or is "just easy" without providing any further explanations.

Comments like “I just don’t get it.” or “It’s too hard and I don’t know why” are representative of this category.

The four categories relating to *rules* are self-explanatory and, in regards to a spectrum, very clear-cut. If a rule is complex, it is difficult, if it is not, it is easy. The same applies to *opacity* and *consistency of rules*. If the use or rule of a feature was predictable and thereby consistent, it was easy, if it was not, it was difficult. A lack of rules or “randomness” or “arbitrariness” were exclusively perceived as a contributor to difficulty.

Tables 2-7 and 2-8 show the distribution of reasons for perceived difficulty/easiness for both, the 100- and 200-level learners. These tables are to show that, for all themes, the distinction between easy and difficult structures moves on a one- or two-dimensional continuum allowing for reliable “if-then” relationships, as described above. They further provide insight into general contributors to easiness and difficulty, which the following analysis will be focusing on. Findings will be presented for both learner groups combined.

Table 2-7: Distribution of reasons *not difficult,* 100-level and 200 learners, in percent

Level	Instruction	Processing Time/ Access.	Cross-Linguistic Transfer	Memori--zation	Cognition	Rule Consis-tency	Rule Complexity	Rule Opacity	Absence of Rules
100	19	9	19	32	2	9	10	n.a	n.a.
200	24	12	15	20	2	17	10	n.a.	n.a.
Mean	*21.5*	*10.5*	*17*	*26*	*2*	*13*	*10*	-	-

Table 2-8: Distribution of reasons *difficult*, 100-level and 200 learners, in percent

Level	Instruction	Processing Time / Access.	Cross-Linguistic Transfer	Memori--zation	Cognition	Rule Consis-tency	Rule Complexity	Rule Opacity	Absence of Rules
100	10	11	14	24	14	7	7	7	6
200	10	11	11	29	7	3	16	4	9
Mean	*10*	*11*	*12.5*	*26.5*	*10.5*	*5*	*11.5*	*5.5*	*7.5*

In general, reasons that were given by students for a particular structure to be perceived as easy are – with a rather high distance to other themes – mainly aspects related to memorization, instruction, and cross-linguistic transfer, with memorization playing a major role with a mean value of 26. The same is true for reasons contributing to difficulty, where memorization is the most frequently mentioned theme as well with a mean value of 26.5. With this value, memorization is the most dominant contributor to difficulty with other themes such as instruction, processing time, cross-linguistic transfer, cognition and rule complexity showing clearly lower values within a very small range between 10.5 and 12.5.

With regard to themes related to rules, it can be said that for easiness, only rule consistency and complexity were mentioned; for difficulty, however, rule consistency, complexity, opacity and absence of rules were themes that emerged.

Overall, the reasons for difficulty seem to be more nuanced than reasons for easiness; in addition, the themes contributing to difficulty are distributed more evenly (with the exception of memorization, which is clearly the most dominant theme) than themes related to easiness.

Discussion

This study aimed at answering the questions as to which grammatical structures are difficult/easy to accurately use in written production and how easy/difficult they are perceived to be by the learners. It was then assessed whether there is a correlation between perceptions and production and what general reasons are for perceiving a structure as easy or difficult. All findings were presented for beginner (100-level) and lower intermediate level (200-level) learners of German as a foreign language.

Overall, the mean accuracy rates as well as perceived difficulty the findings of this study are largely in line with both previous research based on learners'

performance as well as on theoretical considerations on what inherent structural complexity entails (e.g. Housen & Simoens, 2016; DeKeyser, 2005): morphological features, and adjective endings in particular, were demonstrated to show the lowest accuracy level rates and the highest ratings for perceived difficulty. Most of these features require the learner to make a multitude of decisions to arrive at the correct form of the word (Collins et al., 2009).

A comparison of the mean accuracy rates for the two learner groups has shown that significant differences in their performance exist for only four structures, which was surprisingly low: auxiliary choice, dative, determiner agreement and negation. While these groups officially belong to different proficiency levels, institutional conditions need to be kept in mind: between the end of the 100-level and the beginning of the 200-level, there is a long break of more than four months in which many learners have little to no exposure to German. Depending on the point of time in the term, the level of exposure, practice and instruction are not very different for the two levels, which may explain the lack of significant differences in their performance.

The missing association between accuracy and perceived difficulty was surprising since most of the reasons that were provided for a structure to be easy or difficult were formulated against the background of accuracy, that is, learners listed attributes of the features that need to be remembered / accessed / practiced / understood etc. in order to produce accurate output. Consequently, it has to be assumed that learners have very little awareness of the accuracy in their own performance. However, since most of the participants in this study were 100-level learners, the missing language awareness can be assumed to be a result of this early level in the language learning process. Furthermore, the writing prompts are 1) only a snapshot of the learners' performance at the time of data collection and 2) should not be seen as a comprehensive measure since accuracy rates were measured in only

one context, namely text production. In addition, the average occurrence of the features in the writing samples ranged from 5 to 12, which is very high and thereby represents a limitation to the representativeness of the data.

When looking at the quality of the responses for why structures are difficult or easy, the learners overall demonstrated a very thorough understanding of the structures, as was outlined above and is further confirmed when comparing the findings of the present study with Chavez' findings: the number of themes that were identified in the present study is higher than it is in Chavez´ study; it also appears that even the beginner learners had a more nuanced approach to describing difficulties than was accounted for in her study. In this study, nine out of nine identified themes emerged from the 100-level data, while Chavez first-year students only mentioned six out of eight. Surprisingly, the learners in Chavez' study did not mention difficulties that stem from instructional factors, including the amount of practice and explicitness of instruction, which was a very dominant factor in this study. While they do mention the time it takes to access certain features, the theme of intuition is missing. For individual structures, Chavez' learners were mostly concerned with the complexity of rules, which was not the case for the learners in this study, whose perceptions can be considered to be a more nuanced representation of the feature-inherent or objective difficulty.

Another noteworthy difference can be formulated for word order, which was perceived as a difficult feature in the original study but an easy one in this study. Since Chavez included structures that are taught at a higher level, that is, in third year and fourth year, more complex sentence types were included, which could account for this difference.

Yet another difference is the level of language awareness in the learners, as operationalized by their ability to recognize frequencies of a specific feature in either German or English. While Chavez found that "[m]any students were not capable of

assessing the frequency or functionality with which forms occur in either German or English" (p.18), this observation could not be made based on the data collected for this study.

In addition to these differences, there was a high overlap in the identified reasons for why a structure is perceived as easy or difficult. One of the most striking features is *memorization*, which was frequently mentioned in both studies and, in particular, in this study. However, when these findings are compared, it has to be kept in mind that the data elicited in this study was rather limited in comparison to Chavez' data since participants were required to only provide reasons for a total of three structures, that is, one structure per level of difficulty. In Chavez' study, reasons for difficulty were to be provided for all structures in question. Therefore, the findings here provide an insight into general contributors to difficulty and easiness but not into detailed reasons for the difficulty of specific structures.

In Table 2-9, some of the key findings of the present study and Chavez' findings are provided:

Table 2-9: Comparison of findings for difficult features in Chavez (2017) and present study.
Note: Items appear according to the number of occurrences from high to low

Feature / Learner Group	Chavez Study	Present Study
1st year learners	Memorization Complexity of Rules Lack of Experience (=structure had not been taught yet) Lack of Rules/Number of Rules/Speed of Processing	Memorization Cross-Linguistic Transfer / Cognition Processing Time / Accessibility Instruction
2nd year learners	Memorization Complexity of Rules Difference to English Speed of Processing	Memorization Rule Complexity Processing Time /Accessibility /Cross-Linguistic Transfer
Word order	Rule complexity Memorization Lack of Experience Difference to English/Speed of Processing	Not perceived as difficult
Adjective endings	Complexity of rules Speed of processing Number of rules Difference to English/Memorization	Memorization Rule Complexity Processing Time / Accessibility Instruction
Gender / Determiner agreement	Opacity of Rules Lack of Experience Speed of Processing Number of Rules	Memorization Absence of Rules Cross-linguistic Transfer Processing Time / Accessibility
Recognition of frequencies of features in target language	Missing or inaccurate	Mostly accurate
Easiness	Not included	Memorization Instruction Cross-Linguistic Transfer Rule Consistency; continuum
Relationship perceived difficulty and performance	Not included	No correlation identified

Despite the numerous similarities between the findings, there is a major difference in how these findings are contextualized and interpreted. Chavez set out to "[inquire] how learners in four different years of German instruction explain their difficulties" (p.1) by making use of theory and empirical findings as the reference standard to discuss and interpret her findings. For example, the author interprets

answers on what constitutes difficulty from a learner's point of view as an implication of "how they envisioned the learning process itself" and explains that learners "tend toward a mechanistic view of language learning with an emphasis on memorization and a preference for formula-like rules […]" (p.15). These implied learner assumptions are, of course, not in line with current research on what entails and facilitates language learning, which makes her conclude that learners have "unorthodox beliefs" (p.2) about the processes described above. While this holds true and is, for the most part, confirmed by findings in this study, this should not be the only perspective when using such a student-oriented approach, in my view. Although students may not have an accurate understanding of how languages are learned, the consistency in their responses showed that they do have a very thorough and accurate understanding of what constitutes difficulty in regards to individual structures. These findings are in contrast with the conclusion that Chavez (2017) is drawing when she explains that "[student] beliefs about grammar-learning difficulties play out against the backdrop of a growing body of research into what causes these difficulties" (p.2)

She further mentions inconsistencies with regards to the use of terminology such as *memorization*, *rules*, and *exceptions*. As a solution to these inconsistencies, Chavez suggests, as one of her implications, that one of the goals should be that "teacher and students […] in words and practice – share a vision" (p. 19). She summarizes that "[learners] need to realize that their personal theories of language learning are just that – and that research-influenced teaching methods may not accommodate them" (p.19).

While both solutions can be considered as plausible given her findings, using theory and research as a reference norm for assessing student voices reinforces the gap between research and practice by not allowing student-centered research

approaches to complement previous findings or inform future research or instructional decisions.

The findings of this study do not support the inconsistencies identified by Chavez; it has, instead, found the learners to be very reliable and consistent in terms of the insight they have into what constitutes difficulty; with the exception of the connection between their perception of difficulty and their performance, which does, for the most part, not seem to be associated. Furthermore, the nine themes that were identified were very similar to the theoretically and empirically driven literature, whose findings Chavez herself summarizes as

> "the complexity of rules, learners' individual memory capacity, an infrequent or ambiguous occurrence of a form, demands of processing speed, and differences between the L1 and L2 in their respective representation of grammatical features (such as gender)" (p.6)

While Chavez stresses that there is a strong discrepancy between the theoretical understanding of the terms and how they were used by the American students she studied, this could not be confirmed in this Canadian study – with very few exceptions. Students were at times not aware of the difference between linguistic and pedagogical rules and were mostly driven by the latter in their judgement. Overall, however, the mostly shared or congruent understanding demonstrates a very high level of language awareness in the learners.

As was pointed out, the explanation for difficulty was very distinct for each feature and, in most cases, provided an accurate account of the feature's inherent attributes. If we take adjective endings as an example, *memorization* and *rule complexity* are the most frequently mentioned features, followed by *processing time/accessibility* and *instruction*. While these features, to a certain degree, reflect subjective components of difficulty, the answers revealed that the learners are also

aware of what ‘mastering’ the feature entails; knowing whether an adjective is preceded by an article or not, the gender, number and case of the following noun is, indeed, a manifestation of complexity due to the high interdependence of these aspects. Consequently, it takes more time to access this knowledge when producing output. Seeing *memorization* in first place might, at first sight, be considered a misconception of the feature and one might be inclined to say that students are too concerned with memorization. However, after explaining how to find the right ending, what students are left with is a chart that includes 36 endings for three different contexts. Consequently, being overly concerned with memorization might, in fact, not be a misconception on the part of the learner but on the part of how grammar is taught and practiced in class and how it is learned (which is not always the same); which means it may be a result of the fact that students rely more on pedagogical rules as mentioned above. In fact, findings like these might be used to shed light on the relationship between these two types of rules on the one hand and between these rules and their mental representation in the learner on the other (DeKeyser 2005).

The way certain features are taught might be perceived as an overload of information, which leads students to believe that, in order to master this feature, they need to memorize huge amounts of information because it was not presented in *digestible* ways and, thus, does not allow for processing and internalization of the input. Not only does this influence the learners’ perceptions of difficulty, but – in turn – their beliefs about language learning. Grammar does not seem to be paced in a way that facilitates intake; therefore, students should not be blamed for not agreeing with research; while one may conclude that the goal could be to teach students how languages are actually learned, this approach would only be feasible if language learning curricula would literally make room for the inclusion of such processes. Given the density and the resulting time constraints in most post-secondary curricula,

increasing awareness of language learning processes cannot just be added to the existing contents. Findings from studies like this one and the one conducted by Chavez could be a starting point for making decisions on what has to be in the focus if such an approach is implemented.

A look at the five most frequently mentioned themes reveals that they are all, more or less directly, related to or influenced by instructional decisions, which confirms the argument made above; for example: *memorization*: in addition to providing comprehensive tables and overview of what has to be memorized, strategies for how memorization can be facilitated could be increased; *cross-linguistic transfer*: more contrastive (and explicit) approaches could be implemented in the classroom to point out both similarities and differences between the two languages; *processing time/accessibilit*y: exercises that facilitate automatization could be used more frequently and maybe not only outside of the classroom in the form of homework but also during class time; *rule complexity*: often, the complexity of a rule referred to how many decisions a learner has to make in order to arrive at the same form; consequently, learners are in need of tools to process and memorize these decisions; *instruction*: information units for this theme often revolved around whether or not a specific structure is explained well and / or practiced enough.

While Chavez points out the inconsistencies and the learners' lack of awareness (or knowledge) of how languages are actually learned (which is a valid point to make), the findings of this study give reason to believe that such learner-centredness in the research design is – nevertheless – worth-while to be used as a starting point for instructional decisions. Knowing that learners are capable of identifying contributors to difficulties for most structures in question and also knowing that their assumptions about how languages are best learned and taught allows everyone involved in the instructional setting to accommodate the perceptions to a certain degree; either by

facilitating what is perceived as difficult (as long as it is not in conflict with how languages are learned/taught best) or as, Chavez puts it, by helping them realize “that their personal theories of language learning are just that – and that research-influenced teaching methods may not accommodate them” (p.19). In addition, explanations on how languages are learned (best) in specific contexts, as mentioned above, may serve to gap the bridge between student beliefs about these processes and research.

Overall, the learners in this study have proven to possess a high level of readiness as described in the conditioning factors for learner-centredness, as defined by Tudor (1992): *motivation*: students are capable of perceiving needs and understand the learning goals in relation to the features; *experiential traits*: learners demonstrated that they have a high level of prior language learning experience and level of education; *psychological traits:* as was discussed in the description of the learner group, it can be assumed that their aptitude in a bilingual setting like Canada is higher than in other settings and as university students their self-reliance needs to be high as well.

For these reasons, this study argues in favor of instructional approaches that are not exclusively designed from a teacher’s perspective but with a high degree of learner-centeredness. With regard to Chavez study, it can be concluded that – when learners’ voices are involved – one should not only use research or theory to confirm practice. Instead, research and practice should inform each other.

Pedagogical Implications

The fact that the five most frequent themes are related to instruction demonstrates not only the importance of the theme but also the high level of influence that can be exerted on the perception of difficulty. Students described repetition and drills as one of the most effective type of practice. While this request may be read to come from a “backward-looking clientele” (Chavez, 2017, p. 19) – admittedly bringing

back memories of the audiolingual method – it can and should direct one´s attention to frameworks that try to contextualize drill activities in meaningful ways, such as the ACCESS methodology proposed by Gatbonton and Segalowitz (2005). The acronym stands for “Automatization in Communicative Contexts of Essential Speech Segments” (p. 328), which reveals that this approach aims at bringing together CLT and the seemingly contradictory practice of repetition. However, repetition does not necessarily mean that students are required to complete fill-in-the-blanks or other types of mechanical exercises over and over. Instead, repetition could as well entail doing several similar and meaningful communicative activities as practice.

It also speaks to the underlying reasons for students to mention memorization (and also relevant techniques such as mnemonics) so frequently: they are in need of ways to automatize input for it is a central motif and instructional material seems to fall short with regards to facilitating automatization processes.

Another intervention point is the density of the grammatical syllabus, which may lead to the prioritization of drills and memorization techniques. If fewer structures could be practiced in depth instead of aiming at the greatest possible breadth of structures, the students` perceptions may change and the feeling of being overwhelmed may subside to a better understanding of linguistic rules instead of relying on pedagogical ones.

The findings can also be seen as a call for an increase of explicit grammar instruction. The perception of difficulty of a certain feature was shown to be strongly tied to the level of explicit instruction devoted to it. However, the focus should here be on a “focus on form”-approach, which “emphasizes a form-meaning connection and teaches grammar within contexts and through communicative tasks” and not on a “focus on formS”-approach, which is usually detached from meaningful contexts and can be characterized by pattern drills (Brandl, 2008, p. 19). The use of the target

language in grammar instruction should also be implemented more carefully instead of *blindly* following the CLT principle of rich target language input.

Conclusion

This study has replicated the findings from an earlier study by Chavez (2017). Both studies used open questions to elicit data on the reasons for why structures are perceived as difficult or not. With few exceptions, the same themes were identified in the qualitative data. However, unlike Chavez who used theory and research exclusively as a reference norm for assessing the *accuracy* of students' perceptions, which leads to identifying deficiencies in the learners' views, this study focused on identifying the potential of these student voices.

Although the open format of the questions led to qualitative data in which similar themes were identified, one cannot be sure whether the interpretation of the data is congruent with what the students actually wanted to say. Follow-up interviews could serve to validate that the students' output was interpreted correctly. Furthermore, the sample size of this study was rather small, especially for the learners at the 200-level, which reduces the representativeness of the results.

The decision to have learners only provide reasons for the perceived level of difficulty for three and not for all structures was made from a participant-oriented perspective: it was assumed that the time it would have taken to think about reasons for the difficulty of almost 20 structures would have been too long to elicit comprehensive and thorough responses. However, this is – at the same time – a major limitation of this study since it results in an incomplete picture of the reasons underlying the participants' perceptions. For this reason, this study focused on themes relating to general perceptions of easiness and difficulty. Randomizing the order of the features

may have compensated for the thoroughness that may have been missing in the results due to the length of the questionnaire.

While the association between learners' perceptions and accuracy rates was negligible in terms of the number of features that showed significant *p*-values, one should not completely dismiss a potential connection between the two. Not only was the number of significant features higher for the 200-level learners, which may imply that the connection between how difficult a feature is perceived and one's own performance increases with learning progress; the fact that out of two (100-level learners) and three (200-level learners) features with a significant connection there was an overlap for negation may hint at more significant associations when tested with even more structures across different contexts. As was mentioned, the significant correlations are positive, and therefore show a rather unexpected pattern; instead of the expected connection that the higher the accuracy rate the lower the perceived level of difficulty, it is actually the other way around. Future research could examine if more patterns like these can be observed including the reasons.

CHAPTER 3

Study 2: Grammatical Difficulty in German and the Dimensions of Implicit and Explicit Knowledge

This chapter approaches the concept of difficulty from a psycholinguistic perspective. The study that was conducted focuses on the distinction of implicit and explicit knowledge when assessing whether a specific grammatical structure is easy or difficult. Four measures were used to elicit data: the oral imitation and the timed grammaticality judgement test for implicit knowledge and the metalinguistic knowledge test and the untimed grammaticality judgement test for explicit knowledge.

Introduction

For speakers of some languages, like English, German has the notorious reputation of being a difficult language to learn. This difficulty has often been attributed to the morphological richness of the language, which – for example – manifests itself in the form of various noun or article endings depending on distinctions of grammatical gender, case or number. Naturally, morphological richness is a relative concept that largely depends on the morphology of the learner`s own first language (L1) and the typological distance between the L1 and the target language.

The concept of difficulty has been conceptualized and investigated in a number of different ways. Theoretical approaches aim to contribute to an understanding of the L2 acquisition processes (e.g. DeKeyser, 2016, 2005, 2003; Palotti, 2015; Ellis, 2008, 2006) while pedagogically oriented researchers examine difficulty in order to make predictions about the learning process, which are then used to make instructional decisions (e.g. Da Silva & Roehr-Brackin, 2016; Housen, 2014; Spada & Tomita, 2010; Ellis, 2008; Housen et al., 2005; Hu, 2002; Petrovitz, 2001; Robinson, 1996) Despite – or possibly because of – the various approaches and research interests related to the

concept, they all share one characteristic: terminological inconsistency due to a lack of conceptual agreement and comprehensive, systematic knowledge of the concept (Housen & Simoens, 2016; DeKeyser, 2005). In the acquisitional approach, a structure is difficult when it is acquired late (or vice versa for easiness); in instructional settings, difficult structures are usually taught later than easy ones (or vice versa). For example, third-person-singular-s in English is introduced very early in textbooks as it is considered to be easy to learn, but is acquired very late (Ellis, 2006). Thus, the underlying understanding of difficulty is not only inconsistent but even contradictory. Support (and partly an explanation) for this claim is further provided by the fact that naturally occurring and elicited data also seem to show discrepancies when judged according to their difficulty. For that reason, Ellis (2006) claims (and provides evidence) that it does not suffice to differentiate objective or feature-inherent and subjective, that is learner-inherent difficulty, as suggested by DeKeyser (2005). In addition, the concept needs to be approached by examining it based on the different types of knowledge, namely implicit and explicit knowledge.

Since 2006 more studies have therefore started to take this dichotomy into account when approaching the concept, as will be shown in more detail in the review section below. However, so far, most research has focused on English as a second or foreign language. For German, studies on difficulty that take into account the dichotomy of implicit and explicit knowledge do not exist. The present study aims to fill this gap in the research literature. The literature review provides the conceptual background for the study with a brief overview of how scholars have conceptualized difficulty, followed by an examination of the empirical findings relating to the measurement of implicit and explicit knowledge.

Literature review

Defining and conceptualizing the concept of difficulty

As was hinted at above, tackling the concept of difficulty is by itself – ironically – very difficult. Confusing and inconsistent terminology and complexity are two major contributing factors to the inaccessibility of the concept or why it "is so hard to pin down" as DeKeyser (2016) puts it in one of his most recent and most frequently referenced articles on the topic. Housen (2016) summarizes the factors contributing to the complexity of this concept by considering it to be a "multidimensional, multifaceted, and multilayered phenomenon that has cognitive, pedagogical, and linguistic dimensions of developmental and performance aspects, [that] can manifest itself at all levels of language structure, learning, and use." (p. 5). Given this complexity, it does not come as a surprise that, despite the growing body of research on the subject, there is no clearly defined and commonly agreed upon conceptual approach to difficulty. Instead of using theoretical conceptualizations, difficulty is – according to DeKeyser (2003) – often operationalized intuitively, mostly by teachers and (less frequently) the learners.

Collins et al. (2009) as well as Housen (2016) also note the lack of a theoretical conceptualization, but draw on different explanations and examples. In their overview article, Collins et al. (2009) describe the definitions in the acquisitional approach (i.e. definitions that focus on acquisitional orders of certain linguistic features) as being circular in that they state that what is acquired late is difficult and what is difficult is acquired late.

Housen (2016) perceives many of the existing definitions as being descriptive instead of assessing the core question of what constitutes difficulty. In addition, information on difficulty is often inferred information from other studies and research

strands “such as ultimate attainment, learnability and teachability, developmental patterns, order of acquisition, fossilization, [and] effects of instruction” (Housen & Simoens, 2016, p.165). Thus, a theorization of the concept is further avoided but instead a mere conglomerate of bits and pieces is provided.

Given the inconsistencies and deficiencies in the terminology and conceptualization of difficulty, what is – so far – commonly agreed upon? There is an overall agreement that difficulty needs to be approached by considering objective (i.e. feature-inherent) as well as subjective (i.e. learner-inherent) features, which was first thus worded by DeKeyser (2003). Robinson (2001) already discussed this dichotomy but used the term *difficulty* for the subjective variables and *complexity* for the objective ones. Housen and Simoens (2016) refer to this dichotomy as structural (objective) and cognitive (subjective) difficulty. Based on the latter, they provided a taxonomy of L2 difficulty which adds context-related difficulty to the existing dichotomy of learner- versus feature-related difficulty. These context-related factors are referred to as “learning-conditions” (p.164), which are further divided into naturalistic and instructed environments that are either implicit or explicit in nature. Learner-related characteristics can be cognitive or socio-affective (e.g. aptitude, motivation, attitude towards the L1 and the target language, etc.). For this project, feature-related difficulty is the most relevant of all three. Housen and Simoen’s (2016) define “intrinsic feature properties” and “input properties of form-meaning connections” as its two constituting components. The former entails complexity, which can either relate to form or function, while the latter is influenced by “regularity, redundancy, frequency, transparency and saliency” (p. 164). Housen’s and Simoen’s taxonomy is very similar to DeKeyser’s (2016, 2005) approach, in which he describes difficulty as relating to either form, meaning, or the form-meaning-mapping. It merely differs from their taxonomy with regards to the contributing factors for each category and branches. In his approach,

form-related difficulty can result from a “lack of phonological salience […], morphological alternations, or because they are hard to pronounce” (p. 354). Meaning (which is not explicitly addressed in Housen’s and Simoen’s model) in the target language can cause difficulties due to novelty and their degree of abstraction, while form-meaning mappings largely depend on salience and transparency. The latter consists of redundancy (Is the feature in question semantically necessary?), optionality (Is there flexibility in the use of certain features?) and opacity (How easily detectable is the form-meaning relationship; for example “Different forms stand for the same meaning, and the same form stands for different meanings” of the feature in question (DeKeyser, 2003, p.8)).

As mentioned in the introduction, feature-inherent or objective difficulty further needs to be assessed in terms of the different knowledge types as suggested by Ellis in 2006 since (some of) the dimensions identified by DeKeyser (2003) differ depending on whether we are dealing with implicit or explicit knowledge. Aspects that contribute to difficulty related to implicit knowledge are frequency, saliency, regularity and processability. The more frequent/salient/regular/processible a certain feature is, the easier it is to acquire as implicit knowledge; the less frequent etc. the harder it is to acquire. In contrast, the perception of explicit knowledge as easy or difficult is predominantly determined by conceptual clarity and the metalanguage that is required to explain the phenomenon in question (Ellis, 2006).

Since these findings were published, a growing body of research has confirmed Ellis’ call for such an approach. In the following section, the most relevant findings relating to the dichotomy of explicit and implicit knowledge and how they are contextualized in the context of difficulty will be presented.

Implicit and explicit knowledge

The discussion of explicitness and implicitness can refer to three different dimensions, namely learning, knowledge and teaching. Although it may be tempting to assume straightforward, matching relationships between, for example, explicit teaching and explicit learning, Ellis (2009) points out that "there is no necessary correlation between the two pairs of terms" (p. 18), that is, explicitly taught structures may be processed or learned implicitly or vice versa. Research in this field depends strongly on what Ellis (2009) refers to as the "interface issue" (p. 20), namely the assumed relationship or interface between implicit and explicit knowledge and the effect of instruction on both (see Dörnyei, 2009 for a detailed explanation of the interface issue).

Krashen (1981) is the most prominent advocate of the non-interface position, which assumes that implicit and explicit knowledge are entirely separate in terms of how they are stored and accessed. Consequently, it is not possible that one type of knowledge can transfer into the other and (explicit) instruction is considered to be of no use. The strong interface position, with DeKeyser (2007, 2005, 1998) being one of its most famous supporters, claims that "not only can explicit knowledge be derived from implicit knowledge, but also that explicit knowledge can be converted into implicit knowledge through practice" (Ellis, 2009, p. 21). Instruction (explicit or implicit) is therefore of major importance for practice as this is how declarative (explicit) knowledge can be converted into procedural (implicit) knowledge. Finally, the weak interface position can be located between the two aforementioned ones. Generally, it is assumed that a transfer is possible but only under certain circumstances and/or with certain limitations, for example the readiness of the learner. Depending on the perspective within that position, the nature of favored instruction may vary (Ellis, 2009).

Very generally speaking, knowledge refers to "the facts that speakers of a language have learned" (Ellis, 2009, p. 95) and is often described in terms of its accessibility, level of automaticity and consciousness as well as whether or not it can be verbalized:

> Implicit knowledge is the knowledge of a language that is typically manifest in some form of naturally occurring language behavior, such as conversation. It has two major characteristics: it is intuitive and it can be rapidly processed. Explicit knowledge is knowledge about a language. Two types of explicit knowledge can be distinguished. Explicit knowledge in the form of metalanguage consists of knowledge of the technical and semi-technical terms for describing a language.
> Explicit knowledge in the form of analysed knowledge involves an awareness of linguistic form and of form-function mappings which can exist independently of whether learners possess the metalanguage needed to verbalise their knowledge. (Ellis, 2001, p. 252)

Other definitions are less comprehensive but always revolve around the same core idea of conscious versus unconscious processing (e.g. Suzuki & DeKeyser, 2017, 2015; Andringa & Rebuschat, 2015; Rebuschat, 2013; Hulstijn 2005). With regard to the research design of this study, one study of the aforementioned ones is of particular relevance. Suzuki and DeKeyser (2015) wanted to assess “the validity of elicited imitation (EI) as a measure for implicit knowledge” (p.860). While using awareness as a key factor for differentiating implicit and explicit knowledge (see Ellis, 2001 cited above), they add a more fine-grained approach to these two knowledge types by including automatized explicit knowledge which “may be used with little or no awareness” (p.864). Due to the proximity between implicit and automatized explicit knowledge the authors wanted to answer the question whether the EI can be used for implicit knowledge or whether it is, in fact, a measure of automatized explicit knowledge. Based on the finding that scores on measures of metalinguistic knowledge in the study were a “significant predictor of EI” (p. 887), the authors concluded that students were able to rapidly access their language knowledge in the

EI, which can be seen as a confirmation for the assumption that the EI measures automatized explicit knowledge rather than implicit knowledge.

Equally relevant for the present study is the question how these different knowledge types are connected to the concept of (grammatical) difficulty. This connection was first established by Ellis (2006) in the abovementioned article. He hypothesizes that "both types of knowledge are implicated in language proficiency, with different aspects of proficiency (e.g. oral vs. written) drawing variably on the implicit or explicit knowledge of different structures" (p. 432).

Empirical studies of implicit and explicit knowledge

Using four different measures of implicit and explicit knowledge (imitation test, timed and untimed grammaticality judgement test and metalinguistic knowledge test), Ellis (2006) proves the abovementioned hypothesis and concludes that "[t]he difficulty of grammatical structures varied according to whether one is considering implicit or explicit knowledge of the structures. Structures that are easy in terms of implicit knowledge may be difficult in terms of explicit knowledge and vice versa" (p. 459).

Two years later, in 2008, Ellis conducted another study which – among other aspects – provided more evidence for the fact that difficulty needs to be "considered separately for implicit and explicit knowledge" (p. 4). In this study, Ellis contrasted the scores for four grammatical structures of the English language (possessive s, question tags, since/for, 3rd person -s) on measures of explicit (metalinguistic knowledge test) and implicit knowledge (oral imitation test) with the hierarchical processing operations as identified in Processability Theory (Pienemann, 2007). PT assumes that "processing procedures are hierarchical and are mastered one at a time"; "[t]hus, the failure to master a low-level procedure blocks access to higher-level procedures and makes it impossible for the learner to acquire those grammatical features that depend

on them" (p. 8). Processability Theory relies on naturally occurring data and acquisition processes, which are considered to represent implicit knowledge (Ellis, 2006). Ellis found out that while PT predictions were in line with the scores for the measures of implicit knowledge, this could not be observed for the measures of explicit knowledge. Not only did he thus prove that difficulty needs to be considered separately for different knowledge types, but he also showed that "experimentally elicited data can be used to examine interlanguage development (i.e. how learners` implicit knowledge develops) and to make statements and predictions about learners` grammatical proficiency" (p.4)

One year later, in 2009, Roehr and Ganem-Gutierrez confirm his findings but make adjustments to the contributing factors to difficulty as identified by Ellis. For implicit knowledge, they exclude processability but add "communicative redundancy", "opacity of form-meaning mapping" and "opacity of meaning-form mapping" (p. 321). By doing so, they want to increase the applicability of their concept to linguistic constructions and decrease the association to specific theoretical models. For explicit knowledge, they mention "conceptual complexity, technicality of metalanguage, and truth value [as referring] to the characteristics of pedagogical grammar rules and impact on explicit learning difficulty" (p. 321). Pedagogical grammar rules – as opposed to formal grammar rules – "have as their goal to cause someone to produce a language form, that is, getting a learner to perform consistently with regard to some aspect of language behavior" (Seliger, 1979, p. 360).

Dubravac and Takac (2013) looked at Bosnian EFL learners between the ages of 14 and 19. In addition to focusing on subjective difficulty as indicated by the learners, they also focused on the relationship of a structure-inherent aspect of difficulty, namely complexity, and the different types of knowledge. They found that even this structure-inherent aspect of difficulty needs to be considered separately for the two types of knowledge. The participants in this study scored differently for the same structure on

measures of implicit and explicit knowledge (e.g. indefinite articles, plural -s or modal verbs).

Knowing which structures are more or less difficult in terms of the different knowledge types helps optimize the learning environment by implementing instructional decisions, which can accelerate the process of turning explicit into implicit knowledge (as assumed by the strong interface position as outlined above). Knowing more about the former is particularly important for instructional settings, which are usually neglected due to the fact that L2 theories focus on how learners develop implicit and not explicit knowledge. Generally, L2 acquisition is equated with the development of implicit knowledge, while explicit knowledge is connected with classroom instruction and assessment (Ellis, 2006). However, both are of relevance and should therefore be assessed separately when it comes to determining the difficulty of grammatical structures.

Da Silva and Roehr-Brackin (2016) extend the explicit/implicit dichotomy to the concept of rules and thereby to a cross-section of instruction and knowledge: “[p]edagogical rules may be either difficult or easy to learn as explicit, declarative knowledge, but the complexities of the underlying language features themselves relate primarily to the difficulty of learning them as implicit knowledge” (p. 170). The difficulty related to these types of rules is, as has been shown for the different types of knowledge, also determined by different factors. While the difficulty of pedagogical rules is largely determined by “elaborateness”, “conceptual clarity”, “scope”, “reliability” and “truth value”, implicit rules or natural representations “are determined by their objective intrinsic and input-related complexity differentials” (p. 170). Being aware of these differences and the components contributing to both types of knowledge and rules helps make fine-grained and learner-oriented instructional adjustments.

Despite the growing number of studies that focus on difficulty from a psycholinguistic perspective, that is, with a distinction between implicit and explicit knowledge, no such study was found for the German language. Therefore, this project aims at answering the following research questions:

1. Which structures are easy/difficult for English-speaking learners of German on measures of explicit knowledge?
2. Which structures are easy/difficult for English-speaking learners of German on measures of implicit knowledge?
3. Are the same structures easy/difficult for learners at different proficiency levels?

Method

Participants

The study was conducted at a major research university in Western Canada. The participants were adult learners of German as a foreign language at the beginner (100) and intermediate (200-) level. Due to the time-consuming nature of the instruments, the participants were not the same as in Study 1.

Table 3-1 provides an overview of the most relevant biodata information for those students who participated in all necessary rounds of data collection.

Table 3-1: Participant biodata information

Number of participants	**102**
Age span **Average age**	16-62 21.14 yrs
Sex	31 male, 71 female
Level	100 level: 76 participants 200 level: 26 participants
First Languages: English French Chinese Korean Polish French Vietnamese Persian Polish Vietnamese Spanish	 88 2 (incl. bilingual upbringing) 4 1 1 1 1 1 1 1 1
German as L2	30
German as L3 or more	72

The 100-level courses are designed for learners with no or little previous knowledge of the language. However, the coursebook and the teaching approach is tailored to true beginners with no knowledge of the language. The biodata information, however, indicates that the vast majority of learners in the 100-level courses are false beginners who had various levels of exposure to the German language prior to learning it in a formal setting. Only 11% of participants reported that they had not had any exposure to German prior to taking the 100-level course. Exposure mainly consisted of having a friend or family member who “occasionally taught” them “a word or two”.

Research design

Data for the measures of implicit and explicit knowledge was collected in two rounds, using four different instruments based on those developed by Ellis (2006). Implicit knowledge was measured by means of an oral imitation and a timed written grammaticality judgement test; explicit knowledge was elicited by means of a written metalinguistic knowledge test and an untimed version of the same grammaticality judgement test used as a measure of implicit knowledge. This type of test battery was chosen for two reasons: 1) It has successfully been used in previous studies (e.g. Ellis & Loewen, 2007; Ellis, 2006, 2005) and 2) The validity of this approach for implicit knowledge measures has been established by Ellis (2006) and Erlam (2006). Limitations of these measures will be addressed in the discussion section.

Due to the relatively large number of structures in the present study, data was collected in two rounds. In the first round, implicit knowledge was elicited whereas explicit knowledge was measured in round two. Measuring implicit knowledge first ensured that participants had not thought about the grammatical structures in question, thereby drawing on explicit knowledge, which might have influenced the perception of the measures of implicit knowledge. A one-week break was implemented between the two rounds of data collection.

The choice of grammatical structures to be included in the tests was based on the data elicited in Study 1 of this dissertation, in which writing samples were collected, and later judged in terms of their perceived level of difficulty by the learners. However, due to the time-consuming nature of the measures that were used in this study, the number of structures that were in the focus was reduced to 11. The decision to focus on these structures specifically (presented in Table 3-2) was based on several factors: 1) the diversity of the structures, that is, they were supposed to represent different

linguistic domains that are of relevance in the discourse on difficulty referenced above (morphology, syntax and semantics); at the same time, 2) redundancy was to be reduced, which means specific domains should not be over-represented with the same type of structure; e.g. the choice of prepositions vs the choice of two-way prepositions was considered to be too similar in nature. Instead, pairings such as accusative and dative or word order in dependent and in independent clauses were chosen for analysis to further diversify the data.

The decision to limit the relevant structure to 11 was made with the maximum number of test items in mind. For the grammaticality judgement tests, Ellis (2006) recommends the use of 4 items per sentence. Given the time constraints of the setting for data collection and the learner groups, 40-48 sentences were considered to be feasible (which is further in line with literature on the design of these tests (s.Erlam 2006). Therefore, 11 structures were chosen.

The design of all four measures was tested with two pilot groups consisting of 25 and 19 learners of German. These groups were very similar to the study participants with regard to their biodata information. Both groups learned German at the same levels (beginner and intermediate) at another university in the same city.

Table 3-2 provides an overview of the features that were assessed in the different measures:

Table 3-2: Target grammar features for test battery[1]

Feature	Description	Most frequent learner mistake
Subject-verb agreement	Verb endings in German are different for almost each personal pronoun in singular and plural	*Meine Mutter hast (*hat*) ein kleines Auto. My mother has a small car.
Separable prefix-verbs	Some verbs need to be separated when conjugated	*Ich einkaufe oft. / *Ich kaufe oft ein*. I often go shopping.
Negation	While nouns are negated by using *kein/e* (= no), verbs are negated by using *nicht* (not)	*Peter hat nicht (*keine*) Kopfschmerzen. Peter does not have a headache.
Accusative use for direct objects	Direct objects require the use of the accusative; words change their endings based on case	*Marina und Alex kaufen ein (*einen*) Buch. Marina and Alex buy a book.
Dative use for indirect objects	Indirect objects require the use of the dative; words change their ending based on case	*Marina und Alex kaufen meine (*meiner*) Schwester ein Buch. Marina and Alex buy my sister a book.
Verb placement in independent clauses	In independent clauses, the verb needs to be in 2nd position	*Heute ich gehe (*gehe ich*) ins Kino. I`m going to the movie theatre today.
Verb placement in dependent clauses	In dependent clauses, the verb is pushed to the final position in the sentence	*Ich bleibe zuhause, weil ich bin krank (*weil ich krank bin*). I am staying at home because I am sick.
Modal verb placement	When modal verbs are used, they take the 2nd position in the sentence, while the main verb is pushed to final position.	*Jeden Morgen ich kann schlafen bis 8:00 Uhr (*kann ich bis 8:00 Uhr schlafen*). Every morning I can sleep until 8 a.m.
Adverb placement	When a sentence contains a time and a place adverb, the adverb indicating time needs to precede the adverb indicating place.	*Meine Mutter spielt im Wohnzimmer jeden Samstag (*jeden Samstag im Wohnzimmer*) Gitarre. My mother plays the guitar in the living room every Saturday.
Choice of auxiliary in the perfect tense	In German, the perfect tense has two components: an auxiliary verb and a participle; the auxiliary verb is either *haben* (have) or *sein* (be) depending on the action (state versus movement)	*Gestern bin (*habe*) ich ein Buch gelesen. Yesterday I read a book.
Choice of participle in the perfect tense.	In German, the perfect tense has two components: an auxiliary verb and a participle; the participle can contain a “ge” or not, end with a “t” or a “en”; stem vowel changes are also possible and subject to whether the verbs are weak or strong	*Letzte Woche habe ich einen Kleiderschrank gekaufen (*gekauft*). Last week I bought a closet.

[1] The underlined portion is the frequently made mistake; the correct version is provided in italics. The English translation has only been given for the correct version because not all mistakes are transferable to English.

Implicit knowledge measures

Oral imitation test

The oral imitation test was computer-delivered, using OpenSesame, "an open-source, graphical experiment builder for the social sciences" (OpenSesame Website). For the test, two sentences were formulated for each structure and audio recorded by a native speaker of German. One sentence was grammatical, whereas the other one was ungrammatical. Each ungrammatical sentence contained only one mistake. Including more than one mistake would have directed too much attention to the fact that half of the sentences were not grammatical, which would not have allowed for an intuitive response (i.e. imitation) without reflection. For the same reason, the test-takers were not told that the input they were about to hear might contain mistakes.

All sentences contained a proposition that required the students to either agree or disagree. The following sentences are taken from the test and serve as examples for the choice of auxiliary in the past tense. The first one contains no mistake, the second one is wrong with regard to the choice of auxiliary.

> Als Kind bin ich oft nach Deutschland gereist.
> (As a child, I often travelled to Germany.)
>
> *Am Wochenende habe ich ins City Center gegangen.
> (At the weekend, I went to the city centre; should be: "bin ich")

With regard to these two examples, participants were asked to agree or disagree with whether they themselves as children often travelled to Germany (sentence 1) or with going to the City Centre at the weekend (sentence 2). This interim step serves as a distractor designed to prevent students from simply memorizing the sentences instead of reconstructing them (Suzuki & DeKeyser, 2015; Van Moere, 2012). When formulating the sentences for the test, it was also ensured that the

features in question were represented similarly. For example, sentences formulated for the accusative and dative were similar in regards to word order, which means that accusative and dative were used after the verb in all cases, which is usually the third position of a sentence. German would also allow for a use of the object (dative or accusative) in the first (or other) position(s), which was avoided to allow for comparability and to not create unjustified distractors.

OpenSesame allowed for creating the desired sequence by displaying four screens: the first screen included the instructions, the second screen provided the audio input, the third screen asked the students to either agree or disagree by clicking the appropriate button; the fourth screen finally asked them to repeat and record the sentence they had heard in *correct* German. Recordings were also saved by the programme. In order to minimize test-taking effects, such as initial nervousness or fatigue at the end, the items were randomized for each student.

Timed grammaticality judgement test

The timed grammaticality judgement test was delivered in a mixed format, that is, both computer-delivered and in a pen-and-paper format. As previously mentioned, the structures from Study 1 were the basis for formulating grammatical and ungrammatical sentences. Ellis (2006, 2005) suggests using four sentences per structure, two grammatical and two ungrammatical. With a total of 11 structures, the test thus consisted of 44 sentences that varied in length between 12 and 29 syllables.

The abovementioned pilot study provided information about the amount of time each sentence should be shown to the learners. This was important to reduce the likelihood that students would actively reflect upon the sentence they viewed and draw

from their explicit language knowledge. For both the beginner and intermediate level pilot group, the time span was recorded for the first and the last student to respond to each item. The middle value was used to determine the number of seconds that each structure would be shown to the participants of the study. A smaller group of 10 students for each proficiency level served as a pilot group to assess the practicality of these middle values. Based on their performance, small adjustments were made. In the final version of the study design, the allocated response time ranged from 9 to 16 seconds per sentence. Each sentence was shown to the participants as a group on a PowerPoint slide. In the presentation mode, PowerPoint provides the option to set the amount of time for which each slide is displayed before it moves to the next. To reduce fatigue, a 5-minute break was given after the first half of the sentences.

The PowerPoint slide was accompanied by a pen-and-paper form on which the students had four options to choose from. They could decide if the sentence they saw was grammatically “correct”, “incorrect”, they “didn`t know” or they “didn`t have enough time”. The last option was included in order to avoid guessing the correct answer and thereby distort the results. In addition, the students were explicitly asked not to guess in the instructions on the form.

Explicit knowledge measures

Untimed grammaticality judgement test

The untimed grammaticality judgement test was delivered in a pen-and-paper format and used the same sentences as the timed test, which is a common design feature for grammaticality judgement tests (Erlam, 2006). Since one week was in between the two testing rounds and no feedback was provided on whether or not the

previous answers were correct, the use of the same sentences was considered to be justified. Furthermore, this procedure ensures comparability of the findings since the dimension of time is the only variable that is used (and controlled) to differentiate between the elicitation of implicit versus explicit knowledge. Had the items been changed in addition to the variable of time, the items themselves may have contained variables that would have needed to be controlled for. In order to avoid this additional step, the same sentences were used. For this test, the 44 sentences were listed on a test sheet and the students had to decide whether they were grammatically accurate or inaccurate. In case they did not know the answer, they had the option to check “I don`t know”. As was true for the timed version of the test, students were discouraged from guessing.

Metalinguistic knowledge test

The same eleven structures which served as the basis for the other three tests were used for the composition of the metalinguistic knowledge test. For this instrument, however, only one sentence had to be formulated for each structure in order for the desired data to be elicited (Ellis, 2006). This resulted in a total of 11 sentences, all of which contained one mistake. These mistakes were underlined. In the instructions, students were asked to provide an explanation for why the underlined part was wrong. Although an explanation “in grammatical terms” was desirable, students were encouraged to provide an explanation even if they did not remember or know the proper terms. Thus, the intention was to maximize the amount of elicited data. Overall, three options were provided in the questionnaire that participants could choose from: “I do not know why this part of the sentence is underlined, i.e. wrong”, “I know that the underlined part is wrong but I cannot explain why”, “The underlined part is wrong

because…" and "This feature hasn`t been taught yet". The second option targets existing implicit knowledge despite lack of explicit knowledge or a level of explicit knowledge at which they may not be able to verbalize that knowledge (yet). This response option was included in order to ensure that all possible student reactions could be expressed. For the third option, an open format was preferred over a closed one like multiple-choice, for example. Not only would a multiple-choice-format encourage guessing, it would also not allow for a comprehensive and fine-grained understanding of the learners` breadth and depth of metalinguistic knowledge.

Data coding and analysis

For both the timed and the untimed grammaticality judgement test, each item was scored as either correct (1) or incorrect (0). If no answer was given or "I don`t know" was checked, it was counted as incorrect (0) as well. Mean accuracy scores were then calculated for each structure.

The responses from the oral imitation test were transcribed and received a score of either 1 for a correct imitation, or 0 for an incorrect imitation. A score of 0 was also allocated when the answer was avoided. If the target structure was provided correctly but other parts of the sentence were wrong, a score of 1 was allocated and the wrong part was disregarded and may be part of future research. Mean scores were calculated for each structure.

For the metalinguistic knowledge test, a group of 5 experts (i.e. professors and instructors in the field of German as a foreign/second language including both L1 and L2 speakers of German) defined key components of an adequate explanation of each error. This procedure entailed several rounds. In the first round, all group members individually defined components for each structure prior to looking at the data. These

individual findings were then discussed in the group, which resulted in some components being added or removed for reasons such as ambiguity or clarity, or the level of specificity. Prior to the next round, all group members were asked to apply the complete list of identified components to 20 samples taken from the original set of data. Then, everyone provided feedback on their applicability. The goal was to fine-tune the list of components in order to reach the highest level of applicability possible to the actual data-set. In addition to the components which focused on the grammatical subject matter (i.e., the rule or item in question), the acceptable wording for each component or feature was discussed as well. For example, the first sentence contained a modal verb and a main verb, which leads to the main verb being pushed to final position. In this sentence of the MLK, however, the main verb was not pushed to the end but directly followed the modal verb:

> *Studenten an der University of Alberta **können kaufen** auf dem Campus viele Bücher.*
> *(Students at the University of Alberta can buy many books on campus)*
>
> *Correct version would need to be:*
> *Studenten an der University of Alberta **können** auf dem Campus viele Bücher **kaufen.***

In the group, the key components that were identified as adequate and necessary to correctly explain why the sentence in this example was wrong were the following:

"Position of (main) verb is incorrect (since combined with modal verb, which has to be in 2nd position); main verb has to be pushed to the end / final position"

Students were not required or expected to provide more context such as "verb has to be pushed to final position of the sentence" or to mention that the verbs were in second or third position following the subject or the like.

In terms of acceptable terminology, the following were agreed upon:

- Naming the verb that is in the wrong position itself, i.e. saying “kaufen” instead of providing grammatical terminology
- “verb” instead of “modal verb” or “main verb”, if the explanation made it clear which word they were referring to
- “word” instead of “main verb” or “modal verb” if the explanation made it clear which word they were referring to
- examples for modal verbs or verbs in general, if it was clear that they were referring to the correct word

In order for the data to be comparable with the other three instruments, the answers were coded the same way as described above. The decision whether a score of 0 or 1 was reached was based on the question of whether changes based on the explanation that was provided would lead to the correct version of the sentence. The values were then used to compute the mean and standard variation.

In a final round of coding and scoring with a set of 20 different samples (and not the whole dataset), the applicability of these categories was tested and final issues were resolved. This procedure applies only to option (c) on the questionnaire, which required the students to provide an explanation for why the underlined part of the sentence was wrong. The two options “I don`t know why this part of the sentence is wrong” and “I know that the underlined part is wrong but I can`t explain why” received a score of 0. Option (d) “This feature hasn`t been taught yet” was excluded from the analysis because it was checked by only one student.

Results

Explicit knowledge

The first research question asked which structures were easy/difficult for English-speaking learners of German on measures of explicit knowledge. To answer this question, I combined the learners' scores for both measures of explicit knowledge. (Table 3-3). The decision to combine the two scores was based on a dependent t-test that was run to statistically assess the difference between the mean values of these measures. With two exceptions (negation and accusative), the differences were not significant (two-tailed, significance level 0.05). Due to the nature of the instruments, it has to be kept in mind that the mean values for the metalinguistic knowledge test are based on the scores for one sentence per feature and participant whereas the mean for the grammaticality judgement tests is based on four scores per feature and participant. However, since the number of items was decided on based on the nature of the instrument (s. Ellis' 2006 rationale mentioned above), this number (and thus, the number of mean values) is only considered to be relevant for the validity of the instruments, that is, the elicitation of the desired type of knowledge. For the computation (and interpretation) of the mean scores, it was not considered to be of relevance. Since results from both measures (regardless of the number of items) are considered to equally contribute to the representation of elicit knowledge, both means (for the metalinguistic knowledge and the untimed grammaticality judgement test) were considered to have a weight of 50 %. The same rationale applies to the difference in the number of items for the timed grammaticality judgement and the oral imitation test for the elicitation of implicit knowledge.

However, to allow for a more comprehensible overview of the data, the mean scores for all four measures for explicit and implicit knowledge are provided as well. In Table 3-3, scores for the 100-level and 200-level learners are combined; the proficiency variable will be discussed later.

Table 3-3: Mean scores and standard deviation for 11 German structures on two measures of explicit knowledge, 100-level and 200-level learners combined

Feature	Mean UGJT	Mean MLK	Mean Combined (SD)	Standard Deviation
Subject-verb agreement	0.69	0.85	0.77	0.29
Separable prefix-verbs	0.63	0.93	0.78	0.30
Negation	0.66	0.38	0.52	0.28
Accusative use for direct objects	0.47	0.78	0.63	0.32
Dative use for indirect objects	0.59	0.86	0.73	0.32
Verb placement in independent clauses	0.59	0.73	0.66	0.30
Verb placement in dependent clauses	0.80	0.83	0.82	0.27
Modal verb placement	0.75	0.87	0.81	0.25
Adverb placement	0.69	0.88	0.78	0.30
Choice of auxiliary in the perfect tense	0.58	0.59	0.59	0.31
Choice of participle in the perfect tense	0.58	0.39	0.49	0.30

With regard to the concept of difficulty, the cut-off values suggested by Ellis (2008) were applied. For explicit knowledge, structures with a mean score of higher than 0.75 were considered easy; those with a score below 0.50 were difficult. Out of 11 structures, almost half (5), scored above 0.75 (i.e., agreement, separable prefix-verbs, verb placement in dependent clauses, modal verb placement, and adverb placement) and were therefore considered easy. With a value of 0.73, dative is very close to the cut-off score and can therefore be interpreted to also be one of the easier structures. With values of 0.82 and 0.81, verb placement in dependent clauses and modal verb placement are the easiest structures. However, the range between the 0.75

and 0.82 is not very high, which means that within the category *easy* there does not seem to be a relevant difference between the structures of that category in terms of a learners' performance.

The only structure that could be considered difficult was the choice of participle in the perfect tense with a score of 0.49. However, negation, with a value of 0.52, follows and is thus very close to the cut-off score and can therefore be regarded to be a more difficult structure as well.

The remaining structures (accusative, verb placement in independent clauses, and use of auxiliary) range between 0.59 and 0.66, which could be interpreted as moderately difficult.

The standard deviations for the features are all fairly close together with a range from 0.25 to 0.32. It can therefore be assumed that the data was spread out similarly and that there were little to no outliers for the individual features

Implicit knowledge

The second research question asked which structures are easy/difficult for English-speaking learners of German on measures of implicit knowledge. Again, the combined scores for the two implicit knowledge measures are presented in Table 3-4. With the exception of verb placement independent clauses and accusative, the difference between the scores on both measures were not statistically significant, as was true for the explicit measures (dependent t-test, two-tailed, significance level 0.05). Therefore, looking at the combined scores for the two measures was considered to be justified. For the same reasons as provided above, Table 3-4 shows the mean scores for the individual tests in addition to the combined mean.

Table 3-4: Mean scores and standard deviation for 11 German structures on measures of implicit knowledge, 100-level and 200-level learners combined

Feature	Mean TGJT	Mean OI	Mean Combined	Standard Deviation
Subject-verb agreement	0.65	0.69	0.67	0.48
Separable prefix-verbs	0.61	0.46	0.54	0.50
Negation	0.73	0.67	0.70	0.45
Accusative use for direct objects	0.54	0.35	0.45	0.48
Dative use for indirect objects	0.48	0.30	0.39	0.46
Verb placement in independent clauses.	0.59	0.72	0.66	0.48
Verb placement in dependent clauses	0.90	0.63	0.76	0.48
Modal verb placement	0.59	0.71	0.65	0.47
Adverb placement	0.75	0.55	0.65	0.52
Choice of auxiliary in the perfect tense	0.48	0.32	0.40	0.51
Choice of participle in the perfect tense.	0.54	0.56	0.55	0.50

For implicit knowledge, Ellis (2006) suggests a cut-off value of 0.60 for easy structures and 0.45 for difficult ones. Based on these values, 6 out of the 11 structures can be described as easy (agreement, negation, verb placement in dependent and independent clauses, modal verbs, and adverb placement) and three as difficult (dative, accusative and choice of auxiliary).

With regards to the easy structures, most of them have very similar values between 0.65 and 0.70, that is there was – on average – no big difference performance wise and, therefore, difficulty wise.

Only one feature, verb placement in dependent clauses, stands out with a value of 0.76 and is therefore clearly the easiest structure of all. With values of 0.55 and 0.54, choice of participle and separable prefix-verbs fall in the category of moderate difficulty.

As was true for the measures of explicit knowledge, the standard deviations for the measures for implicit knowledge are close together as well with a range from 0.45 to 0.52. Therefore, it can be concluded that the data was spread out similarly for each structure. However, while the standard deviations are close together for implicit knowledge, they are higher as they are for explicit knowledge, which means that the data is, overall, spread out more than is the case for explicit knowledge.

In addition to relying on the overall mean values to compare differences in the performance on the four measures, as was done by Ellis 2006, a two-tailed dependent t-test was used here to statistically assess the difference between the means. In Table 3-5, the mean values for both knowledge types are presented, including the difference between the means as well as *t*- and *p*-values of the dependent t-test (significance level 0.05).

Table 3-5: Difference between explicit and implicit scores for 11 German grammatical structures, all participants combined

	Explicit Mean (SD)	**Implicit Mean (SD)**	**Difference between means**	***t*-value**	***p*-value**
Subject-verb agreement	0.77	0.67	0.1	**-3.8**	**.001**
Separable prefix-verbs	0.78	0.54	0.24	**-3.9**	**.001**
Negation	0.52	0.70	- 0.18	**9.2**	**.000**
Accusative use for direct objects	0.63	0.45	0.18	-1.5	.144
Dative use for indirect objects	0.73	0.39	0.34	**-11.4**	**.001**
Verb placement in independent clauses	0.66	0.66	0.0	**-3.8**	**.003**
Verb placement in dependent clauses	0.82	0.76	0.16	**-3.1**	**.001**
Modal verb placement	0.81	0.65	0.16	**-7.4**	**.000**
Adverb placement	0.78	0.65	0.13	0.1	.915
Choice of auxiliary in the perfect tense	0.59	0.40	0.19	**-2.1**	**0.04**
Choice of participle in the perfect tense.	0.49	0.55	- 0.06	**3.9**	**.002**

With the exception of accusative and adverb placement, the results from the dependent t-test were significant (in bold), which means that the Null Hypothesis that the means are equal can be rejected. Consequently, the performance on the measures of implicit knowledge and explicit knowledge was significantly different for most features.

Overall, the mean scores on the measures of explicit knowledge are – for 9 out 11 structures – higher than those for the measures of implicit knowledge. The only two exceptions, indicated by the negative values, are negation and choice of participle.

When looking at which structures were easy or difficult according to the cut-off dates provided by Ellis (2006)(see above), only one overlap could be identified for difficult structures: choice of auxiliary is among the three most difficult structures for both measures. For easy structures, a higher overlap could be observed and consequently more consistency. With one exception (separable prefix-verbs), all of the structures that are defined as easy on measures of explicit knowledge fall into the same category for implicit knowledge, namely modal verb placement, verb placement in dependent clauses, adverb placement and subject-verb agreement. It is notable that three out of these four structures revolve around the concept of placement.

As previously mentioned, the overall higher values for explicit knowledges do not apply to two structures, namely negation and choice of participle. Not only is the score for negation higher on measures of implicit knowledge than it is for explicit knowledge, but the mean of 0.70 is the second highest (and thereby almost easiest) structure on measures of implicit knowledge. Remarkably, it is the second most difficult structure on measures of explicit knowledge with a mean of 0.52. For choice of participle, the difference is even lower with a value of 0.06. As for the level of difficulty within the context of values for implicit knowledge, a mean of 0.55 is located roughly in the middle of the range of differences from 0.39 to 0.76.

Table 3-6 provides a summary of the findings by dividing the values into the different levels of difficulty, namely *easy*, *moderate* and *difficult*, for all structures; categories are given for all participants combined.

Table 3-6: Results categorized in different levels of difficulty, all participants combined

Feature	Explicit	Implicit
Subject-verb agreement	Easy	Easy
Separable prefix-verbs	Easy	Moderate
Negation	Moderate	Easy
Accusative use for direct objects	Moderate	Difficult
Dative use for indirect objects	Moderate	Difficult
Verb placement in independent clauses.	Moderate	Easy
Verb placement in dependent clauses	Easy	Easy
Modal verb placement	Easy	Easy
Adverb placement	Easy	Easy
Choice of auxiliary in the perfect tense	Moderate	Difficult
Choice of participle in the perfect tense	Difficult	Moderate

Despite the statistical assessment that the mean values for the two measures are significantly different, there is an overlap in terms of the categories of difficulty for four features (agreement, placement in dependent clauses and for modal verbs as well as adverb placement). This inconsistency results from the fact that the different levels of difficulty are based on the mean values and cut-off dates as “somewhat arbitrarily” suggested by Ellis (2006, p. 460). However, due to the statistical analysis, the information can be added that the four features that are marked as easy on both measures are significantly easier on measures of explicit knowledge than they are on measures of implicit knowledge.

Overall, on measures for implicit knowledge, six features were easy while that only holds true for five structures for explicit knowledge. Inconsistency could be observed for the remaining seven structures of which three were more difficult on measures of explicit knowledge (negation, placement in independent clauses and

choice of participle) and four on measures of implicit knowledge (separable prefix-verbs, accusative, dative and choice of auxiliary).

Comparison between proficiency levels

To address the third research question about the patterns of ease or difficulty for learners at different proficiency levels, comparisons were made between the mean values for implicit and explicit knowledge for students enrolled in the 100 level German course to those enrolled in the 200-level course. The comparison of 100-level and 200-level students' test performance provides information about the development of learners' linguistic knowledge from a cross-sectional research design perspective.

A two-tailed independent t-test was used to assess the differences between the means (significance-level 0.05). Table 3-7 displays the results for both learner groups on measures of explicit and implicit knowledge.

Table 3-7: Means, t- and p-values for 100- and 200-level learners for implicit and explicit knowledge

	Explicit Knowledge				Implicit Knowledge			
Feature	**Mean 100-level**	**Mean 200-level**	***t*-value**	***p*-value**	**Mean 100-level**	**Mean 200-level**	***t*-value**	***p*-value**
Subject-verb agreement	0.71	0.82	-1	.310	0.57	0.77	**2.8**	**.003**
Separable prefix-verbs	0.75	0.80	0.5	.603	0.42	0.65	1.5	.138
Negation	0.44	0.60	-1.2	.223	0.64	0.75	**2.7**	**.004**
Accusative use for direct objects	0.55	0.71	-0.7	.502	0.40	0.49	1.8	.072
Dative use for indirect objects	0.71	0.74	-0.1	.952	0.36	0.42	-1.3	.213
Verb placement in independent clauses	0.63	0.69	-1.9	.656	0.62	0.69	1.2	.221
Verb placement in dependent clauses	0.78	0.85	-1.4	.168	0.76	0.76	1.1	.240
Modal verb placement	0.82	0.79	0.5	.603	0.58	0.72	1.3	.0201
Adverb placement	0.77	0.79	-0.9	.325	0.63	0.67	-0.9	.354
Choice of auxiliary in the perfect tense	0.62	0.57	0.7	.445	0.31	0.49	-0.1	.909
Choice of participle in the perfect tense	0.34	0.63	-1.3	.199	0.50	0.60	1.9	.064

With the exception of agreement and negation (in bold), the results from the independent t-test were not significant, which means that the Null Hypothesis that the means are equal has to be confirmed. Consequently, there is no significant difference in the performance of the 100-level and the 200-level learners on the two measures.

As was done for the overall difference between measures of implicit and explicit knowledge, the mean scores and cut-off values suggested by Ellis (2006) were used here as well to add another dimension to the findings. Table 3-8 presents the different levels of difficulty for both measures and learner groups.

Table 3-8: Results categorized in different levels of difficulty, 100-level and 200-level learners

	Explicit Knowledge		Implicit Knowledge	
Feature	**100-level**	**200-level**	**100-level**	**200-level**
Subject-verb agreement	Moderate	Easy	Moderate	Easy
Separable prefix-verbs	Easy	Easy	Difficult	Easy
Negation	Difficult	Moderate	Easy	Easy
Accusative use for direct objects	Moderate	Moderate	Difficult	Moderate
Dative use for indirect objects	Moderate	Moderate	Difficult	Difficult
Verb placement in independent clauses.	Moderate	Moderate	Easy	Easy
Verb placement in dependent clauses	Easy	Easy	Easy	Easy
Modal verb placement	Easy	Easy	Moderate	Easy
Adverb placement	Easy	Easy	Easy	Easy
Choice of auxiliary in the perfect tense	Moderate	Moderate	Difficult	Moderate
Choice of participle in the perfect tense	Difficult	Moderate	Moderate	Easy

In Table 3-8, the levels of difficulty for 100- and 200-level learners, as expressed in the form of mean values from the four measures of implicit and explicit knowledge and interpreted based on the cut-off values suggested by Ellis, are different for agreement, negation, and choice of participle on measures of explicit knowledge.

On measures of implicit knowledge, a difference in category can be observed for six structures, namely agreement, separable prefix-verbs, accusative, modal verb placement, choice of auxiliary and choice of participle.

Lower levels of difficulty for the 200-level learners can be observed for only three out of the eleven features for explicit knowledge (agreement, negation, and choice of participle). For implicit knowledge, however, there was a decrease in difficulty (i.e. a change from either “difficult” to “moderate” or from “moderate” to “easy”) for six out of the eleven structures for the 200-level learners, namely agreement, separable prefix-

verbs, accusative, modal verb placement as well as choice of auxiliary and participle in the perfect tense. Agreement and participle choice are the only two features that were less difficult for 200-level learners on measures of both explicit and implicit knowledge. However, these findings always need to be assessed against the background of 1) the cut-off values suggested by Ellis (2006) on the one hand and 2) the statistical difference between the mean values on the other.

Discussion

Overall, the fact that no significant difference exists between the 100- and 200-level learners was – at first sight – surprising since higher scores would have been expected for higher proficiency levels. However, all four measures were very different from the forms of assessment that the learners were familiar with from the classroom setting. Thus, it can be assumed that the novelty of the measures may have contributed to negating differences in performances that could have been observed with more familiar forms of assessment, that is with forms of assessment for which 200-level learners may have had higher levels of test-taking skills, which may positively influence their performance (Dodeen 2008). Despite the lack of a significant difference between the two proficiency levels, an overall of the values for six out of eleven structures for implicit knowledge from the 100- to the 200-level learners could be observed, which means that difficulty levels seem to have decreased. This supports the claim that achieving implicit language knowledge takes longer than is true for explicit knowledge.

The fact that the average scores for measures of implicit and explicit knowledge were different confirms findings from previous studies by Ellis (2006) and thus provides further support for the claim that difficulty needs to be approached with this distinction in mind.

Overall, consistency was different for the two measures. For easy structures, there was an overlap for four out of eleven structures, while no such overlap existed for difficult or moderately difficult structures. Given the fact that different cognitive processes are involved when performing these tests and that aspects that contribute to difficulty related to implicit knowledge are different than those contributing to explicit knowledge (s. DeKeyser, 2003), it was further not surprising that inconsistent patterns could be identified in the profiles for easy and difficult structures.

The fact that six structures were easy on implicit measures but only five on measures of explicit knowledge was, at first sight, rather unexpected if one contextualizes these findings within the strong interface position where it is stated that "[…] explicit knowledge can be converted into implicit knowledge through practice" (Ellis, 2009, p. 21), implying that implicit knowledge is what is more difficult to attain. However, the fact that the study was conducted with anglophone speakers in an Inner Circle country with very limited grammar instruction in K-12 schooling, may account for the performance on explicit measures to be – relatively – lower than on measures for implicit knowledge. In addition, the type of instruction the learners receive, which does not focus on explicit instruction, may also be reflected in the results.

When discussing the reasons for the differences in the scores, which are largely in line with both theory and research, as has been stated above, the different aspects that contribute to difficulty related to implicit and explicit knowledge need to be reviewed again, namely frequency, saliency, regularity and processibility for implicit knowledge and conceptual clarity and metalanguage for explicit knowledge (Ellis, 2006). For easy structures, there was an overlap for four out of the eleven structures, which were categorized as easy on both measures: agreement and placement in dependent clauses, modal verbs and adverb placement. Consequently, it can be

assumed that the aspects mentioned above as contributors to difficulty related to implicit and explicit knowledge were equally met.

With regards to implicit knowledge, frequency and regularity have to be pointed out in this context. Agreement and verb placement are omni-present in the classroom from the very first lesson and there are no exceptions to the concept of agreement and to verb placement in dependent clauses; therefore, it is not surprising that these structures are easy on both measures.

Unlike for easy structures, less consistency, or overlap, could be identified for difficult ones. For no feature could an overlap between the categories of difficulty be identified. Accusative and dative were categorized as difficult for implicit knowledge but only moderately difficult for explicit knowledge. While both are fairly frequent as direct and indirect objects, saliency can be a problem in settings other than the classroom. For example, articles are often contracted when German speakers have conversations in natural settings: instead of using “einen” or “eine” (indefinite article, accusative), one often hears short versions such as “nen” or “ne”, which reduces saliency drastically. However, communicative saliency is thus not affected because learners should be able to determine which object is direct and which is indirect from context. For example, hearing “Ich gebe der Freundin ein Buch” (= I give the book to a friend) makes it clear that the book was given to the friend and not the other way around.

In addition, German morphology is not only a problem for learners of German but often for L1 speakers of German as well. Accusative, dative (and genitive) endings are often confused and used wrong (Sick, 2004), which poses a challenge in terms of regularity of a form. Also, in terms of processability, finding the correct case ending

requires numerous mental processes and knowledge of number, gender and the question whether an article is definite or indefinite or used at all.

Another finding was the categorization of the choice of auxiliary for implicit knowledge as difficult and, at the same time, the choice of participle as moderate. For explicit knowledge, choice of auxiliary is moderate whereas choice of participle is difficult. While they are distinct grammatical structures that follow their own rules and regularities, they have to be used together in German in order to form the past tense (Perfekt). When being compared, the English present perfect and the German Perfekt are very similar form- wise. They are both compound tenses consisting of an auxiliary verb as well as the third form (i.e. the participle) of the main verb. While English does not require a choice between two auxiliary verbs, German learners need to decide whether to use to have (*haben*) or to be (*sein*). This item is rule-based, that is, *sein* is used for verbs that indicate movement (e.g., to go – gehen; to swim – schwimmen, to travel – reisen) as well as those that describe a change of state (e.g., to fall asleep – einschlafen; to die – sterben; to wake up – aufwachen). The past tense of verbs that do not fall into these categories are (usually) formed with *haben* as the auxiliary. For the participle, it does not suffice to judge a verb based on whether or not it indicates movement or a change of state. In order to arrive at the correct form of the participle, learners need to understand the difference between weak and strong verbs in order to decide whether the participle form has to end in “-t” (weak and mixed verbs) or “-en” (strong verbs). Usually (!), the prefix “ge” is added to most verbs but the list of exceptions to the rules described above is extensive. It therefore does not come as a surprise that the participle is difficult on measures of explicit knowledge since it requires more metalanguage to be understood as well as a very clear representation of the concept of strong and weak verbs (conceptual clarity). Alternatively, higher scores on measures of implicit knowledge may simply be a result of memorization. Instead of

going through all the steps to arrive at the correct form as described above, learners may simply memorize the individual forms.

For the choice of the correct auxiliary, such knowledge is not necessary because verbs of movement and change of state are very relatable. Here, components adding to difficulty on measures of implicit knowledge, such as frequency, saliency and regularity, seem to be more relevant. While the past tense is used quite frequently in German, saliency and regularity may be an issue for choice of auxiliary in natural language settings. Often, the forms of the auxiliary verb are contracted (as was true for articles mentioned above), e.g. “hasse” or “bisse” instead of “hast du” or “bist du” or they are left out: “Schonmal dort gewesen? (Have you ever been there?) instead of “Bist du schonmal dort gewesen”?

So far, findings have been discussed based on the assumption that the four measures actually did measure implicit and explicit knowledge. Referring back to the study by Suzuki and DeKeyser from 2015, in which they have demonstrated that the Oral Imitation Task in their study did, in fact, not measure implicit but automatized explicit knowledge, one may wonder if that holds true for this study as well. While this question can only be addressed hypothetically and not empirically, some major differences between Suzuki and DeKeyser’s study and this one exist, which might lead to a valid conclusion: 1. Learners in the study by Suzuki and DeKeyser learned Japanese as a second and not a foreign language, as was the case in this study, 2. the mean average length of instruction in Japanese as about 39 months at the time of the study, which is much higher than the length of instruction for the learners of this study, which is approximately 6.5 (100-level learners) to 13 months (200-level learners). In regards to both 1) and 2), the level of exposure was much higher for the learners in Suzuki and DeKeyser’s study than was true for learners in the present

study, which means that there was more time for the Japanese learners than for the German learners to automatize explicit knowledge. 3. Learners in this study were not explicitly told that the sentences they are about to hear contain a mistake. They were only asked to repeat the sentence they heard in “correct German”. Suzuki and DeKeyser (2015) have referred to the amount of exposure, the length of instruction as well as to the fact that they explicitly told the participants that they need to correct the sentence as reasons for their conclusion that automatized explicit knowledge was tested and not implicit knowledge. Due to the conditions of this study, however, there is no reason to assume that the same applies to these findings.

Other studies have focused on the validity of timed and untimed grammaticality judgement tests as measures for implicit and explicit knowledge. While numerous studies have confirmed their validity – which has been the reason for this study to rely on these two measures – (e.g. Ellis & Loewen, 2007; Erlam, 2006; Ellis, 2005) , more recent and fine-grained research has hypothesized and confirmed that “GJTs are too coarse to be measures of [implicit knowledge], and that the different types of GJTs measure different levels of [explicit knowledge]” (Vafaee et al., 2016, p.1). “Type” of GJTS refers to both, the dimension of time restriction as well as the question whether the sentences that are to be assessed contain mistakes or not. Gutiérrez (2013) has found that for responding to grammatical stimuli, implicit knowledge is used whereas ungrammatical stimuli lead to accessing explicit knowledge.

Another more general issue with timed GJTs is that asking learners to judge the grammaticality of a sentence may lead to misleading data since this does not show whether learners know the reason for why a certain sentence is wrong or what the correct version of the sentence would have to be unless they are asked to provide the correct answer for the sentences that they identified as ungrammatical.

Limitations

Like every study, this study has its limitations. In addition to the issues related to the individual measures – as described in the previous section – a larger sample size, especially for the 200-level learners, would have led to more representative results. However, since learner groups are usually much smaller in the higher levels than they are in the beginner levels, larger institutions would have to be chosen as a research setting.

The coding processes used for the metalinguistic knowledge test may be considered to be subjective and not stringent. However, by having five people involved in the process, subjectivity was reduced to a large extent. Instead of formulating core elements for the explanations within a group, textbook explanations may have been used to supplement these explanations. Since the categories were applicable to the data, this step was not deemed to be necessary. In addition, interpreting student answers in open format questions always leaves room for misunderstandings. While this is an issue that is a result of the very nature of open questions, follow-up interviews may have reduced the risk of such misunderstandings.

In response to the issues regarding the Oral Imitation Test identified by Suzuki and DeKeyser (2015) as well as research on the validity of the Grammatical Judgement Test (Vafaee et al., 2016; Gutiérrez 2013), the results from the four measures for implicit and explicit knowledge of this study may benefit from being supported by findings on different measures for implicit and explicit knowledge. While the choice to use these measures in particular was based on the study conducted by Ellis (2006) and research on these measures by Erlam (2006), which have thoroughly established

their validity, the more recent concerns described above need to be kept in mind when designing future research on this topic.

Pedagogical implications and future research

This study has supported existing research on the connection between different knowledge types and the concept of difficulty for the German as a foreign language classroom. Despite the fact that the tasks participants carried out are not communicative in nature and do not resemble naturally occurring speech or communicative tasks that should be part of communicative oriented or task-based classrooms, the findings are still considered to be transferrable to the classroom setting for several reasons: 1. Ellis (2008) demonstrated that implicit knowledge, despite being elicited in the form of instruments used in experimental settings, such as the ones used in this study, resembles naturally occurring speech as manifest in hierarchical processing operations identified in Processability Theory (Pienemann, 2007). 2. Not all forms of assessment in the classroom are purely communicative in nature or represent authentic communicative settings. Therefore, it is considered to be legitimate to conclude that assessing difficulty should not only be approached based on this distinction of implicit and explicit knowledge in research, it may also be beneficial to take this distinction into account for textbook design and other instructional decisions.

The examples that were discussed in the discussion section show that language instruction can benefit to a large extent from focusing on the distinction between implicit and explicit language knowledge and how it is connected with the concept of difficulty. However, since research that focuses on the distinction between implicit and explicit knowledge in the context of difficulty is still relatively new and underrepresented, instructional settings as well as material such as textbooks have not approached

grammatical structures with this distinction in mind. The fact that the two components from a compound tense involve different kinds of learning processes suggests the need to modify the usual 'one-size-fits-all' approach to grammar instruction. It further implies that future research is needed (for the German language) that focuses on more structures with the goal of depicting a relatively complete assessment of (the most frequently used) structures in the (German) language (classroom).

For both learners and teachers alike, being able to identify which structures are easy or difficult to learn either implicitly or/and explicitly, can result in a more efficient work environment since the amount of exposure to a certain structure could be adjusted accordingly with “explicit leaning […] allow[ing] for one-trial learning as well as […] minimal exposure to input”, and the outcome of implicit learning “[being] dependent on ample exposure to input” (Da Silva & Roehr-Brackin, 2016, p. 320). Yet, the very general and fundamental question whether it is easy or difficult structures that should be targeted in institutionalized settings has been subject to controversy for the past 30 years (Da Silva & Roehr-Brackin, 2016).

In addition to the very concept that (grammar) instruction (including textbook and curriculum design) should take the differences for implicit and explicit knowledge into account, these insights can be used as a reflective tool for instructional settings. The contributing factors to implicit and explicit knowledge can serve as a map for instructional fine tuning. For example, if features appear to be difficult to learn as explicit knowledge, it is likely that instruction should address the clarity of the metalanguage that is being used to introduce or talk about a certain feature. Alternatively, if that same feature is easy to acquire as implicit knowledge, there may not be any need to present it explicitly.

Likewise, difficulty of mastering structures on measures of implicit knowledge can be a hint that the representation of the feature in question needs to be modified, namely its frequency, saliency or regularity. If the learners' explicit knowledge is overall higher than implicit knowledge, for example, it is very likely that the learning environment needs to implement more opportunities for students to practice.

However, given the high time pressure that many teachers are facing due to very dense curricular requirements, such an approach is only feasible once certain criteria are met: More comprehensive research is needed in order to find out whether 1) there are regularities in the findings across institutions, proficiency levels etc. and 2) what factors might exist that lead to a certain representation of a grammatical feature (e.g. the use of specific textbooks, teaching style or the teacher's attitudes and beliefs about grammar instruction). Finally, 3) the gap between research and practice might be reduced if such forms of assessment would be designed by the researchers for the classroom (and not exclusively for research).

To conclude, when supported by research and implemented successfully, such an approach would not only be beneficial for the learners and their learning outcome, but it may also serve as a measure for instructional quality insurance by adding another way of operationalizing the learning environment.

Chapter 4

Study 3: Metalanguage in German as a Foreign Language Textbooks as a Source of Learning Difficulty

Chapter 4 focuses on the use of metalanguage in six beginner textbooks of German as a foreign language and how it can be a source of learning difficulty. Metalanguage is approached quantitatively by analysing the overall vs. the distinct number of metalinguistic terms, the number of explained vs. assumed terms as well as the number of transparent vs. opaque terms.

Introduction

Despite the rise of the internet and other forms of digital media that offer limitless access to authentic input in the target language, the textbook, usually a collection of adapted material, is still dominant in institutional language learning; be it at the level of curriculum, programme or lesson planning (Guerrettaz & Johnson, 2013; Tomlinson, 2012; Maijala, 2010). Consequently, the textbook constitutes a major component in the teacher – student – instructional material – triad; all three of which can be considered to mutually influence each other. In addition to being in the centre of instructional decisions, the textbook also, either implicitly or explicitly, reflects and influences – *inter alii* – both the language and the learner in question. For example, the choice of grammatical structures included in a textbook for German as a foreign language and how the structures are sequenced implies an understanding of the German language on different levels: Which structures are relevant and why? Which structures are difficult and why? When should which structure be introduced and why? At the same time, assumptions about the learner are made: What do they already know about the language in question? What will they perceive as difficult and how can these perceptions be addressed? What is of interest to the learner? What type of exercise

will they find helpful? This paper explores implicit assumptions about what is difficult about learning German through an in-depth analysis of six GFL textbooks.

For speakers of some languages, German has the reputation of being a notoriously difficult language to learn. One of the first (humorous, non-scientific) elaborations on the difficulty of the German language for native speakers of English was written by Mark Twain in 1880. Entitled, “The Awful German language”, Twain’s essay predominantly focuses on the morphological richness of German, which manifests itself in concepts such as case, gender, inflection of adjectives, or verb conjugation, and separable prefix-verbs. One of the most recent popular scientific contributions on the reputation of German being a difficult language to learn (or use) was made by Sick (2016, 2004). In his books, he documents how even L1 speakers of German are often unable to use cases correctly; for example, the dative is frequently used instead of the genitive or accusatives are used where a dative is required. Google searches on the question whether German is a difficult language to learn complement the popular scientific notion of the language’s difficulty. In more scientific contexts, however, difficulty is not a simple concept to identify.

In applied linguistics, it has been approached in a variety of ways, ranging from the acquisition perspective, to the linguistic, the psycholinguistic and the pedagogical (Collins et al., 2009). For this project, the latter is of particular relevance. In the center of the pedagogical perspective is complexity, which is – unlike in other perspectives such as the linguistic one – not considered to be a feature-inherent construct; instead, it is defined in relationship to the “the explanation for a given feature” (ibid, p. 339). This connection, in turn, revolves around the amenability of a rule to explanation. Factors which influence this amenability are a) the number of exceptions to a rule, b) the amount of metalanguage necessary to understand and/or explain the feature or

the arbitrariness of a specific feature (Collins et al. 2009, and Ellis, 2006 on characteristics of learning difficulty as explicit knowledge) and c) the vagueness of the metalanguage (ibid. 2009).

As a morphologically rich language, German requires the use of technical metalanguage for concepts such as case, for example dative or accusative, when explained to L1 speakers of languages that are less morphologically complex and may not possess similar morphological qualities. In institutional language learning settings, this metalanguage is predominantly found in textbooks, namely in above mentioned "explanation[s] for a given feature", which is why the pedagogical approach to difficulty has been chosen for this study.

Without the understanding of these concepts, and with it, the associated metalanguage, it will be difficult (albeit not impossible) for learners (in institutionalized settings) to produce accurate output. If the learner's L1 does not have equivalents, this metalanguage can be problematic.

Based on the understanding of pedagogical difficulty as explained by Collins et al. (2009), this paper presents the results of an analysis of GFL textbooks focussing on the metalanguage used in grammar explanations.

Literature review

As mentioned above, metalanguage is directly connected with the pedagogical understanding of difficulty. According to this understanding, complexity is what makes a specific feature easy or difficult. However, this complexity does not manifest itself in the nature of the very feature but in the pedagogical rules that are used to explain them. The explanations, in turn, usually rely on the use of metalanguage, especially in institutionalized language learning settings, be it in the form of teacher explanations or in explanatory (grammar) sections in textbooks that are used in the classroom. Thus, a difference in terms of difficulty can exist between a structure itself and the way it is learned. For example, adding an s to the third-person-singular for the present tense is a rather simple rule in English but it is still difficult for the learners to learn and correctly apply that rule (Ellis, 2006) Consequently, when approaching metalanguage from the pedagogical perspective, metalanguage should be further assessed in the different contexts it is used, namely the L2 classroom and instructional materials and also in relation to how it is connected with learning processes and outcomes in the target language.

What is metalanguage?

When approaching the concept of metalanguage, one needs to be aware of its position in regards to the distinction of two different types of linguistic knowledge in the field of Second Language Acquisition. Generally speaking, producing output in a foreign or second language requires a different type of knowledge than performing tasks that involve metalinguistic analysis using metalanguage. In the literature, this distinction comes with various labels depending on whether the focus lies on types of knowledge (Anderson, 1983; Bialystok, 1978), processing (Mc Laughlin, 1978),

developing (Krashen, 1978) or using (Cummins, 1991; Felix, 1981; Lamendella, 1977) the knowledge (see White & Ranta, 2002 for an overview). The use of metalanguage is here considered to be a manifestation of a certain type of knowledge, namely "explicit" (Bialystok, 1978) or "declarative" (Anderson, 1983) knowledge. The description of linguistic facts is (with very few exceptions) impossible without some kind of specialist terminology and so metalanguage can be considered to be an integral (but not always mandatory) part of explicit knowledge (see Ellis, 2009).

As was mentioned, metalanguage can also be described as "declarative" in nature. Just like explicit knowledge, the fact that it can be articulated by the (language) learner is of major importance here. Declarative knowledge is often defined as factual, or "that-knowledge" in contrast to procedural or "how-knowledge". According to Andersen (1983), declarative knowledge is less gradual and easier to achieve than procedural knowledge. While the former can be seen as a simple encoding of a cognitive unit, the latter can only be achieved by "doing" (p. 215), for example in the form of practice. Here, metalanguage or metalingual terminology are a manifestation of verbalizable *facts* about the language in question, namely that-knowledge.

In this paper, metalanguage and metalingual terminology are used interchangeably and they are both simply understood as "language about language" (Johnson & Johnson, 1998, p. 212) or "talk about language", that is the discursive activity that has language as a topic. Thus, they include all linguistic domains, but my focus is on the language used to explain grammatical structures in particular. As such, metalanguage can vary in terms of its "technicality" (James & Garrett, 1992), depending on the nature of the feature in question. Technicality refers to the fact that a term is part of a very specific field (of study) and is, therefore, part of a specific register. When technicality is mentioned in the context of metalanguage, the register

that technicality refers to is the linguistic register. Technicality is often directly identifiable because the terms are not part of everyday language. For specific words, however, they can only be recognized in the specific context they are used in as a technical term because they are also part of everyday language (usually with a different meaning) as the examples below demonstrate. When we look at the concept of case, we can break it down, but it will be impossible to completely avoid technicality. As a starting point, the following definition of *case* is taken from the GFL textbook *Kontakte* (2012):

> "German speakers use a case system (nominative for the subject, accusative for the direct object, and so on) to indicate a function of a particular noun in a sentence."
> (p. 20)

This *function* refers to whether a certain *noun* (phrase) is, for example, a *direct* or *indirect object,* which means breaking the concept down entails the use of even more technicality. When looking for definitions of these terms, and thus breaking down the concept even more, the following can be found:

> "The direct object is the receiver of the action and is in the accusative case" (*Denk Mal*, 2012, p.22)

> "In the English sentence: *The boy asks the father,* the DIRECT OBJECT of the sentence is *the father.* He is being asked; he is the target of the verb's action."

While the degree of technicality does decrease in these definitions, in contrast to the initial one for the concept of case, they also start to become less specific: while the directness of an object is explained, the very idea of *object* is not.

How can metalanguage be classified?

Berry (2010) offers another distinction with respect to metalanguage, namely "scientific" and "pedagogical" (p. 34) terminology. The former is characterized as *distinct* in that there is a clear and recognizable difference between metalinguistic terminology and regular vocabulary and *precise* in that there is no ambiguity in understanding the term since it refers to only one concept (Berry, 2010, p. 34). For German, Kleineidam (1986) refers to this ambiguity as *terminological equivalence,* meaning that different terms can refer to the same concept and are, therefore, equivalent or synonymous to each other. On a very general level, some German grammatical concepts can either be described with terminology that is closer to Latin or closer to contemporary German. For example, instead of *Kasus* (case) one can say *Fall* (case); for the concept of comparison, either *Komparativ* or *Steigerungsform* can be used. For the past perfect, one may find either *Plusquamperfekt* or *Vorvergangenheit* (pre-past). While the Latin words are clearly more *scientific*, and *distinct* and *precise* (Berry, 2010), the German words are context-sensitive. *Fall* (case), for example, can only be recognized as a grammatical term when used in a grammatical context. In everyday language, the word *Fall* can refer to a *legal case*, a particular situation (*in her case*) or even to *downfall*. However, while these words are homonyms, they do not always go back to the same root, that is, their etymology is different.

Kleineidam adds the idea of *terminological divergence*, which refers to differences in terminology based "either on the underlying understanding of acquisitional processes or on structural differences" (p. 173).

Beyond these broad categorizations, metalanguage can be described in terms of its 1) transparency vs. opaqueness and 2) its iconicity (Berry, 2010). The understanding of a transparent term is facilitated by the fact that its denotation as such "gives a clue to the meaning" (p.46) while for opaque terms, "there is no obvious relationship between the term and its referent" (p.50). Iconic terms are a direct reflection of the referent (p. 45). An example that is frequently mentioned for English is the *-ing*-form. In English, for example, when referring to the progressive form as the "ing-form", or to conditional sentences as "if-clauses", we have a physical resemblance between the signified and the signifier.

Transparency literally refers to how a learner can *see through* a term to understand what is meant by it. For example, when comparing the terms *verb* and *action word*, the latter usually makes it easier for the learner to understand, or predict, what is meant by the term because it unmistakably establishes the connection between the word and its meaning; it is therefore clear, or transparent, which is not true for the Latin term *verb*. Based on this explanation, the term *action word* would be considered as transparent and the term *verb* as opaque. However, while the level of transparency could thus be increased, the level of accuracy may decrease. Not all verbs are action words, which may lead to a limited or wrong understanding of the term.

In this study, another distinction is made, namely the difference between metalanguage that is *assumed* by authors to be part of learners' existing metalinguistic knowledge or whether it is *explained*. For the concept of case, for example, an explanation may look as follows:

> German speakers use a case system (nominative for subject, accusative for the direct object, and so on) to indicate the function of a particular noun.
>
> (*Kontakte,* 2012, p. 20)

In this example, the explanation of *case* requires the use of more metalinguistic terms (i.e., *function* and *noun*), which are not explained. An alternative approach is to limit the use of additional grammatical terminology when introducing a metalinguistic concept (i.e. *article*) that may not be familiar to students:

> *Articles are words such as the, a, and an, which precede nouns.*

The dichotomy of explained versus assumed is also part of Heringer and Keller-Bauer's (1984) approach to determining the difficulty of metalanguage, which is based on the following three criteria: 1) comprehensibility *(Verständlichkeit*) 2) connectivity (*Anschlussfähigkeit*) and 3) adaptivity (*Anpassung*). Comprehensibility is reached when a specific metalinguistic term is explained or defined to ensure that the learner has a clear, distinct and comprehensive understanding of its meaning. Consequently, uncommented use of metalinguistic terms contributes to an increased level of difficulty for the learners. This distinction is congruent with the understanding of assumed versus explained grammar in this section.

The second criterion, connectivity, refers to terminological consistency across languages, which allows students to *connect* metalinguistic knowledge they have in one language with that of another. Connectivity is of particular relevance for GFL textbooks that are designed for a specific learner group, for example L1 speakers of English. The following example serves to demonstrate the explicit use of terminological connectivity:

> Both English and German have a small group of MODAL AUXILIARY VERBS that modify the meaning of another verb. Modal verbs express ideas such as permission, ability, necessity, obligation, or desire to do something.
>
> (*Wie geht's*?, 2007, p. 136)

The concept of adaptivity, criterion 3, implies the degree of appropriateness for the target group, that is to what extend are the explanations *adapted* to learners' needs. These needs can relate to aspects such as age group, learning objectives or the learners' L1 (Ivancic, 2010). For example, the extent to which metalanguage is used depends on the context. In an immersive language learning environment consisting of children, explicit instruction and – automatically, the use of metalanguage – is practically non-existent. In contrast, post-secondary language-learning environments that targets adult learners usually includes more explicit instruction and metalanguage.

Metalanguage in L2 instruction

Research on the use of metalanguage in instructional settings generally revolves around 1) identifying the metalinguistic terminology or knowledge that learners (or teachers) possess (e.g. Gebhard et al., 2014; Gánem-Gutiérrez & Roehr-Brackin, 2011; Berry 2009, 2004, 1997; Cajkler & Hislam 2002; Andrews 1998), 2) the role it plays in the foreign language classroom (e.g. Arnawa, 2017; Hu, 2011; Roehr et al., 2009; Macken-Horarik, 2009, 2008; Roehr, 2006; Elder & Manwaring, 2004; Serrano, 2001), and 3) how this knowledge correlates with learning outcomes, other forms of knowledge or language awareness (e.g. Gutiérrez, 2013, 2012; Hu, 2011, 2002; Roehr, 2008; Elder et al., 2004; Renou, 2001; Keen, 1997; Widdowson, 1997). Often associated with this research is a prescriptive discourse on the practical question of which role metalinguistic terminology or knowledge *should* play in instruction. This, in turn, usually leads to the controversy around the dichotomy of explicit versus implicit teaching (for overviews see Kang, Sok & Han, 2018 or Norris & Ortega, 2000). Metalanguage is associated with the former, namely explicit teaching. Explicit instruction serves to direct the learner's attention to linguistic properties of the target

language, which are presented explicitly, that is, features are explicitly described and rules are explained.

In addition to the approaches to metalanguage described above, the use of metalanguage in the language learning classroom can cause difficulties for various reasons: L2 learners who have not been confronted with metalanguage in their L1, or those who have acquired a second language in natural or immersion settings and have therefore not been confronted with metalanguage, may – to a certain extent – be overwhelmed by the novel analytical approach.

Furthermore, even if learners are familiar with the use of metalanguage in their L1, a lack of transferability of concepts from one language to another may lead to problems.

Metalanguage in L2 learning and in German instructional materials

Currently, SLA research offers relatively limited evidence in support of a direct connection between the difficulty of metalanguage and L2 learning outcomes. The connection between metalinguistic terminology (as one component of explicit knowledge) and learning difficulty was first established by Ellis (2006), but has since rarely been in the focus of research. He measured difficulty in terms of learners' performance on measures of implicit and explicit knowledge for 17 grammatical structures of English. The group of 229 learners from different L1 backgrounds displayed different degrees of difficulty depending on whether a student's implicit or explicit knowledge was tested. Thus, third-person-singular-s and indefinite articles were *easy* in terms of explicit knowledge but difficult in terms of implicit knowledge.

Shiu (2011) builds on Ellis' findings and provides further support for his conclusion that difficulty operationalized as scores on performance tests depends on knowledge types. Shiu found that "learners' perceptions of grammatical difficulty vary according to their implicit/explicit knowledge of the features in question" (p.5). Furthermore, the easier a rule could be articulated by learners, the easier it was perceived to be.

One of the few empirical studies of German as a L2 was conducted by Ritterbusch et al. (2006). These researchers correlated learners' self-assessed expertise of *English* metalanguage with their accuracy in using German cases. They found that "[p]oor understanding of grammar correlated with weak performance on non-canonical items" (p. 37). Canonical vs. non-canonical was here understood in terms of frequency of its use. However, metalanguage in Ritterbusch et al. (2006) was operationalized in the form of learners' assessment of their metalinguistic knowledge as indicated on a Likert scale and not as the actual metalanguage that was used to describe or explain case distinctions in German. Additionally, students were asked to assess their knowledge of English and not German metalanguage. While this approach offers valuable insights into the learners' linguistic analytic ability, which is most likely connected to their ability to understand and make use of metalingual instruction in German, it is indirect in nature and does not explicitly focus on German metalanguage

Grün Johansson (2018) analysed grammar explanations in three textbooks for German as a foreign language that are used in secondary schools in Sweden. In these explanations, she identified references to specific models of language acquisition such as the cognitive approach or the audiolingual method. The most frequently underlying approaches were the grammar translation method as well as the natural approach in all three textbooks. Her research supports findings from Funk (1995) who describes

the use of metalanguage in textbooks for German as a foreign language as directly dependent on the underlying understanding of how a language is learned. However, according to him, these views started to become less dogmatic, meaning different understandings of how a language is acquired started to co-exist in the 1980s. Due to and as a direct reflection of this variety, the canon of metalingual terminology increased as well and was no longer the reflection of one specific understanding of meta-language.

Ivancic (2010) used the framework suggested by Kleineidam (1986) to examine grammars of German for learners of Italian and found a high degree of divergence and inconsistencies, which led her to conclude that teacher training (be it in the context of primary, secondary or post-secondary education) should entail instruction on how to deal with metalanguage in both grammars and textbooks.

Engel (1995) analysed the use of metalanguage in textbooks based on their 1) linguistic accuracy or precision (i. e., the congruency and consistency of the meta-language with the underlying conceptualization of grammar), 2) its consistency within and across textbooks, and 3) its usefulness for an instructional environment. He formulated a long list of problems or shortcomings. For example, the terms *Imperfekt* and *Präteritum* (= *simple past tense*) are used interchangeably in several textbooks without explicitly indicating that they both refer to the same structure and thus imply that there may be a difference in meaning. Overall, he identified a lack of linguistic precision and consistency and assessed most approaches as not suitable for GFL classrooms.

Other approaches by scholars have focused on individual structures and their metalinguistic representation in either textbooks or grammars and are often theoretical or prescriptive in nature, that is, they make suggestions for metalanguage that *should*

be used to represent the structure in question. Diewald and Kresic (2010), for example, have suggested a two-field scheme ("Zwei-Felder-Schema") to describe the meaning of different modal particles (such as *denn* or *doch)* in German as a foreign language. They suggest providing the language learner with two different pieces of information for each particle: 1) How does the speaker perceive the immediate context, that is, what does he/she think? and 2) What does the speaker actually say? The aim is to make the process of choosing a specific modal particle transparent. This is particularly relevant since many modal particles in German are not translatable to English. To this day, such transparent and multi-dimensional use of metalanguage can rarely be found in GFL textbooks.

From this review of the literature, we see that there has been little empirical research on the topic of difficulty relating to German as a foreign or second language and that the relationship between specific attributes of metalanguage and related concepts of difficulties has not been addressed. Consequently, this study aims to begin to fill in this gap in the research by examining what metalingual terminology is used to explain grammatical structures in GFL textbooks. Specifically, the analysis of textbooks addressed the following research question:

What kind of metalanguage is used in the explanation of thirteen grammatical features in bilingual beginner textbooks for German as a foreign language?

Method

Textbook selection and description

The textbooks for this study were partly selected based on an unpublished survey amongst GFL university level teachers in Canada, conducted by a professor of German at a major research university in Western Canada (K. Misfeldt). Teachers from 21 Canadian universities participated in this survey. Overall, 9 different textbooks were mentioned. Generally, these textbooks can be divided into monolingual, or global, textbooks and bilingual textbooks. Monolingual or global textbooks are not designed with a specific learner group in mind in terms of their L1, which means the entire textbook is in German, including grammatical explanations and metalanguage. In order to allow for valid findings, only bilingual textbooks targeting L1 English speakers were chosen from the survey for analysis since translating German metalanguage in monolingual textbooks to English would have been considered to change the original source text too much. Furthermore, this procedure would have led to comparing translated metalanguage with original English metalanguage, which was not considered as valid. Based on these criteria, six textbooks named in the survey qualified for analysis. One of them, however, was not accessible at the time of data collection and was thereby replaced by a textbook that equally met the criteria mentioned above and had been known and used by myself.

Finally, the following six bilingual textbooks were selected. They were all designed for beginner learners of German:

- *Berliner Platz 1 Neu (Lemke et al., 2009)*
- *Treffpunkt Deutsch (Gongolweski et al., 2013)*
- *Wie geht's? (Sevin & Sevin, 2007)*
- *Sag Mal (Anton et al., 2014)*

- *Deutsch – Na klar! (Di Donate & Clyde, 2015)*
- *Kontakte (Tschirner et al., 2012)*

Each textbook covers all four skills, namely reading, writing, speaking, and listening. However, although not explicitly stated, writing exercises are overall lower in number than speaking, listening and reading exercises. This circumstance can most likely be attributed to the fact that all textbooks target beginner learners of the language. In all books, the units are structured thematically: they revolve around a certain topic, for example “Who am I and what do I like do to?”, “Talents, Plans, Obligations”, (Kontakte) “Families, Countries, Languages” (Wie geht’s?) or “Leisure Time” (Treffpunkt Deutsch). For these topics, relevant vocabulary and grammatical structures are usually presented separately in each unit, which means learners are first confronted with input such as texts, dialogues, images etc. and then grammar and vocabulary are presented in individual sections that can usually be found at the end of the unit. With almost no exception, learners are confronted with the sequence presentation – practice – production (PPP) in each unit. The textbook *Kontakte* is using its very own sequence, which starts with a presentation phase and continues with the “receptive recall” in which learners have to recognize contents but are not yet required to produce output on their own. Instead, they are then asked to repeat what the teacher says in the “choral response” phase. Finally, in the production phase, they are addressed individually and need to produce their own output; for example, they need to answer a question. While this sequence refers to speaking only, the PPP sequence can predominantly be found in all six textbooks and for every skill.

In one of the six textbooks, it is explicitly stated that a communicative approach is pursued in and with the textbook (*Kontakte*); in the other five, the introductory sections explain that (everyday) communication and conversation is in the centre of

each unit without the Communicative Approach being mentioned explicitly. References to other approaches, such as TBLT, cannot be found.

Data collection and coding

The question of what kind of metalanguage was used was narrowed down to 13 of the grammatical features that were targeted in Studies 1 and 2 of this dissertation. These structures are as follows:

1. Word order dependent clauses
2. Word order independent clauses
3. Adverb placement (time before place)
4. Separable prefix-verbs
5. Accusative case with direct objects
6. Dative case with indirect objects
7. Negation of verbs (nicht)
8. Negation of nouns (kein)
9. Modal verb choice
10. Modal verb placement
11. Choice of auxiliary in the past tense
12. Choice of participle in the past tense
13. Agreement

Since 11 out of these 13 features were the same as in Study 2, the selection criteria were the same as well: diversity in terms of linguistic domain with a focus on morphology, syntax and semantics, which are often mentioned in the literature on linguistic difficulty, as well as the complexity the feature entails. Since the concept of negation is broken down into negation of verbs and negation of nouns in textbooks, this distinction was added here as well. Modal verb choice was further added to have

another feature representing the domain of semantics (which would otherwise be under-represented).

The term *feature* is understood as “the elements into which linguistic units, such as words, can be broken down”. These features can further be divided into morphosyntactic features (such as gender, number, case), morphosemantic features (such as tense, aspect, mood) and morphological features (such as inflectional class, stem). In the textbooks, the term *feature* is often replaced by *structure* and refers to either words (morphology) or sentences (syntax).

In order to identify which metalanguage was used to describe these features, the textbooks were searched for the sections in which grammatical explanations were provided. Both, the table of contents and the index and/or glossary for each textbook were used to identify all relevant occurrences. In almost all cases, grammar features were presented in a special grammar section, which was either integrated within a unit or, more frequently, placed at the very end.

The research question *what kind* of metalanguage was used to explain the different features refers to different aspects, which naturally leads to differences in both the coding and the analysis of the data for each aspect:

1. Total number of metalinguistic terms used for a feature
2. Number of distinct metalinguistic terms used for a feature
3. Number of explained versus assumed metalinguistic terms
4. Number of transparent versus opaque terms

Since the selection of the textbooks was based on the survey results and not, for example, based on similarities in the quality, amount and presentation of metalanguage in the textbooks, they were all different with regard to these aspects. Therefore, the quantitative dichotomies mentioned above were considered as a way of

operationalising the metalanguage in a way that applied to all textbooks and could therefore be examined.

The following examples demonstrate how types and token were identified for the different aspects.

Total vs. distinct number of metalinguistic terms

In the following example, an explanation for the concept of dependent clauses is provided that can be found in one of the textbooks that were analysed for this study:

> […], the first clause is the main clause. The clause introduced by a conjunction is called a dependant clause. In German, the verb in a dependent clause occurs at the end of the clause.
>
> (*Kontakte,* 2012, p. 135)

In this explanation, we have a total of eight metalinguistic terms (underlined). The compounds “main clause” and “dependent clause” were counted as one token because they form a conceptual unit. As far as the type *distinct terms* is concerned, we have 5 tokens in this example:

clause, main clause, dependent clause, conjunction, and verb.

Here, too, conceptual units are taken into account as such: although “clause” appears three times, it is considered as distinct in combination with these modifiers.

For some grammatical features, the explanations provided information on more than one characteristic of that form, usually in later chapters of the textbook in question. For example, separable prefix-verbs were first introduced for use in independent clauses:

> As you know, the prefix of a separable-prefix verb occurs at the end of an independent clause.
>
> […]

Afterwards (sometimes in the same grammar chapter, other times in later chapters of the textbook) additional information was provided, for example on how to deal with separable prefix-verbs in dependent clauses:

> Where there are two verbs in a dependent clause, such as a modal verb and an infinitive, the modal verb comes last, following the infinitive.
>
> (*Kontakte*, 2012, p. 137)

Since the choice of structures that were analysed in this study was based on the structures examined in Studies 1 and 2, only those aspects about these structures were in the focus that were of relevance in the previous studies as well. For example, in Studies 1 and 2, separable prefix-verbs were only examined when they appeared in independent clauses in the present tense, that is, when they are first introduced. Later occurrences, such as separable prefix-verbs in dependent clauses (example 2 above) or with other tenses, were not of interest. Therefore, only metalanguage that was used in these contexts (usually the first mention of the feature) was analysed in this study.

With regard to the two examples above, only the first mention of the metalinguistic terms in the first example (independent clauses, example 1) was analyzed. Including other occurrences of the structures would not have allowed for a comparison of the findings.

Number of explained vs. assumed metalinguistic terms

For counting the tokens for this type, the analysis had to go beyond looking at the metalinguistic features in a specific grammar unit, mentioned above. A term was considered to be *assumed* to be part of the learner's pre-existing linguistic knowledge when it was used in a sentence without being explained, exemplified or in the form of other options such as signal grammar; the latter being a form of input enhancement (Sharwood Smith, 1993) that serves to increase the visibility, or salience, of grammatically relevant aspects, such as word endings, by presenting them in a different color, font type or supported by highlights.

The decision on whether a term was assumed or explained only made sense when not only the grammatical explanations for the relevant features were taken into account but the entire textbook up to the point in which the explanation that was analyzed occurred. For example, when the term *object* is used to explain the concept of accusative in chapter 4 of a book, the analysis of whether it is an assumed or explained term cannot only be based on the explanation of the term (or lack thereof) in the grammar section of chapter 4. Instead, all grammar sections of chapters 1 to 3 have to be searched for the term *object* in order for the term to be labelled as *assumed* or *explained*. When it was not previously mentioned up to chapter 4 and chapter 4 itself also did not provide an explanation, the term was considered to be *assumed*. When, either in chapters 1-3 or in chapter 4, an explanation was provided, it was coded as *explained*. Therefore, when all relevant terms in the explanations for the features in question were determined, they were searched for in the previous chapters up to the point where the structure was presented, either manually by browsing through each individual chapter and / or by using the glossary.

Since only beginner level textbooks were part of the analysis, this procedure was considered to be valid as no mentioning of a certain term could have been part of a preceding textbook for a lower level of German.

Number of iconic, transparent and opaque terms

After the terms of each feature had been identified and counted, they were analysed for the types *iconicity* and *transparency*. These categories are based on Berry (2010) and describe the relationship between "le signifiant": the signifier or the word that is used, and the underlying meaning or concept: the "signifié" or signified (Saussure 1976).

Iconicity is easy to identify since the form itself resembles the underlying concept; for example the expression "denn-Satz" can be used to describe sub-clauses introduced by the conjunction *denn* (= because) in German; "if-clause" in English refers to conditional sentences introduced by "if", or, by using the expression "ing-form" instead of "progressive-form", a directly recognizable connection between form and concept is established.

The difference between opacity and transparency, however, is more difficult to determine as it is more complex and multi-faceted. Generally, transparency (or opacity) is determined by the directness or closeness between the signifier and the signified. For example, when looking at the synonyms *suffix* and *ending*, one can identify a direct connection between the term *ending* and what it is referring to. This does not hold true for *suffix*, which does not directly *give away* what it may mean; at least when one does not have a knowledge of the Latin language. A student who is familiar with Latin, and / or the etymology of the word, in cases where the meaning of a word has not changed

much over time, is likely to know that *suffix* contains *sub*, which can mean *behind* (amongst other meanings) and is therefore close to *ending*. In other words, when approaching these terms with a background in Latin or in linguistics, most of these terms have a direct connection between signifier and signified and are, therefore, transparent: terms such as adverb, pronoun or prefix directly refer to what the word does: an adverb is an addition/specification to a verb, *ad verb(um)*; a pronoun replaces a noun, that is, it stands for it – *pro noun*, and so on.

However, for this transparency to be valid, learners would need to know Latin, or at least the translation or etymology of these terms. Since this is not assumed to be the case, these terms were counted as tokens for the feature opaque.

Based on these examples, one may be tempted to assume that less technical metalanguage is transparent while more technical terms are opaque. However, while this holds true for these examples (and for most Latin expressions) that is not always the case: the terms *strong* or *weak* (verbs), for example, are not technical terms at first sight; yet, their meaning in the grammatical or linguistic register is not directly transferrable from their form. In German, the distinction between *weak* and *strong* verbs refers to how the infinitive, simple past and past participles of verbs are formed. In particular, the difference consists in the fact whether or not the stem vowel of the verb in questions changes (strong verbs) or not (weak verbs). The terminology weak and strong was introduced by Jakob Grimm (1819), who referred to *strength* as the "word's ability to form its temporal stem out of itself" (Frieske et al., 2004, translation from German provided by author). Thus, *weak* and *strong* were coded as opaque. The same is true for *case*; while it exists in laymen speech, its meaning in the linguistic register is different and therefore not implied either. A lack of technicality usually only contributes to the transparency of a term when the meaning of the words in laymen

speech and in the linguistic register are identical or, at least, very similar (s. verbal *bracket* or simple *past*).

Transparency and opacity are seen as opposite ends of one spectrum instead of a dichotomy. To account for the different degrees of transparency or opacity, the labels *pseudo-transparency* and *pseudo-opacity* were introduced and used for coding the data, which is described in more detail in the following section.

Pseudo-transparency and pseudo-opacity

One component of transparency is the similarity between layman speech and linguistic register, when the term(s) in question exist in both. Naturally, when the understanding is identical, and the connection between signifier and signified is direct, the term can be considered as transparent. However, referring to the congruence between laymen speech and linguistic register does not suffice when other aspects are involved that are necessary to fully understand the meaning of a term. In the following examples, the meanings of the underlined terms are identical, or very similarly, in layman speech and the linguistic register:

direct vs. indirect object
regular vs. irregular verb
dependent vs. independent clause

However, all of these three dichotomies share that they are not absolute in nature but relative. Regularity, dependence and directness can only be judged when the reference norm is known. In order to understand regularity of a verb, the learner needs to know that regularity refers to the inflection, or – more precisely – the conjugation of the verb and whether or not it follows the typical patterns of the language in question.

The same is true for the concept of “dependency”; as trivial as it may sound, it needs to be known that dependence of a sentence is manifested in whether it can stand alone or not both semantically (is the thought complete) and syntactically (is the sentence complete).

The directness of objects is measured or assessed based on its connection with the verb of the sentence in question. A direct (or accusative) object is what the subject is directly acting upon, whereas the indirect (or dative) is linked to the subject *via* the direct object and is therefore considered to the be recipient of the action, that is, the indirect object. If these reference norms are not known or vague, clarity or transparency of the term is affected. Therefore, and in order to distinguish them from the ones that are unmistakably transparent, these terms were labelled “pseudo-transparent”. The prefix “pseudo” is chosen to cover the fact that “at first sight” these terms may appear to be transparent, due to their overlap in meanings in different registers, when – in fact – one needs to know the reference form. An alternative would have been “almost transparent”; however, this would not have accounted for the fact that the term may need to be assumed or perceived as transparent, but only the fact that *something* is missing from complete transparency. “Alleged” or “assumed” would have been more viable alternatives but “pseudo” was preferred due to it being more ‘catchy’. In addition to pseudo-transparent terms implying a reference norm, this category was also used for terms that, just like the ones above, imply an “at first sight” understanding of the meaning. For example, while “sentence” and “phrase” are likely to be immediately understood as a conceptual unit due to the high (alleged) overlap between their laymen understanding and their linguistic counterpart, awareness might be lacking that these units are semantically and grammatically distinct. In contrast, a pseudo-opaque term is one that seems opaque at first sight but – in fact – is not. For

example, the verbal aspect *progressive* seems to be opaque but is transparent once the connection to “action is progress” is made.

All tokens for the types described above were entered into the software Excel. Reliability in coding was assured by having two university teachers (one for English and one for German who were both familiar with both languages) identify (most of) the tokens for the types described above in addition to myself. Both teachers were trained in using the coding criteria: after an introductory explanation of the coding rules, they each coded two random samples from the data to apply the coding criteria. These rounds provided the opportunity to ensure that they had understood the rules correctly and also served as a chance to answer questions they had. After these trial runs, 75 % of the data was coded by all three, that is, myself and the two teachers. For the 75% of the data that was coded together, agreement was reached for about 95% of the tokens (after discussing them). The same procedure was done for the types *assumed* versus *explained*. An overview of all coding categories including examples can be found in Appendix F and an explanation for how multi-word expressions were coded in Appendix G.

Results

Total number and number of distinct metalinguistic terms and explained vs. assumed terminology in each textbook

Table 4-1 presents the results of the metalanguage analysis for each textbook, that is, the total number of occurrences and the number of unique terms, including the ratio between these two and the proportion of assumed versus explained metalanguage.

In addition to how a term was generally identified as assumed or explained as explained above - the ratio of assumed and explained terms in Table 4-1 refers to the use of the *distinct* terms in the explanations for the 13 structures. For *Berliner Platz,* for example, it was checked how many times the 14 distinct terms that were identified were used as either explained or assumed terms in the explanations of the 13 structures in question.

Table 4-1: Metalinguistic terms in six German as a Foreign Language textbooks

Textbook	***Total # of terms***	***Distinct Terms total***	***Ratio***	***Assumed (%)***	***Explained (%)***
Berliner Platz 1 Neu	17	14	1.2	79	21
Treffpunkt Deutsch	47	34	1.4	23	77
Wie geht's?	67	44	1.5	25	75
Sag Mal	74	44	1.7	19	81
Kontakte	55	34	1.6	29	71
Deutsch – Na klar!	58	38	1.5	41	59
Range	57 74-17	30 44-14	0.5 1.7-1.2	60 79-19	60 81-21
Mean	53	34.7	1.48	36	64

At first sight, the textbook *Berliner Platz* can be considered as an outlier for both the total and distinct number of terms. For the remaining textbooks, the range between the highest and the lowest number of total metalinguistic terms is higher than the range between the highest and lowest values for the distinct number of terms, 27 and 10, respectively. What these numbers tell us is that the overall *verbosity* of the explanation for individual features varies more than the *diversity* with which grammatical features are explained. Verbosity is understood here as the ratio between the total and distinct number of terms, or – in other words – a measure to show the relationship between the overall and distinct number of words used. For example, the value of 1.7 (for the textbook *Sag Mal*) indicates that the total number of words that were used to explain a feature was 70% higher than the number of distinct words. When looking at this number in terms of real numbers, that means that for 10 distinct terms a total of 17 other terms was used.

With a value of 1.2, *Berliner Platz* has the lowest value while *Sag Mal* has the highest one with 1.7. The remaining textbooks have values at or around 1.5 with a range from 1.4 (*Treffpunkt Deutsch*) to 1.6 (*Kontakte*). Based on the understanding that verbosity can be seen as repetition, it can be said that for *Berliner Platz*, there was hardly any repetition of words (with a value of 1.2), that is metalinguistic terms in the explanations were mostly mentioned just once.

For the ratio of assumed versus explained knowledge it can be said that, with the exception of *Berliner Platz*, all textbooks have a clearly higher value for explained terms than they do for assumed terms. With a value of 59% for explained terms, *Deutsch – Na klar!* has the lowest proportion, while *Sag Mal* has the highest value with 81%. When looking at the average values, it can be said that roughly two thirds of the terms are explained (64%), while one third is assumed (36%).

With regard to this dichotomy, the analysis further showed that many of the explained terms were not explained when they first occurred but later on in other chapters.

Total and distinct number of terms and assumed vs. explained terminology per feature

In Table 4-2, results are presented for each grammatical feature. In addition to the total number of metalinguistic terms and the distinct ones (including their ratio), average values as well as the range for each feature are provided. Again, the percentages for assumed versus explained terminology are given including their range and mean values; as was true for the ratio for assumed and explained terms in the six textbooks (Table 4-1), the percentages here refer to the distinct number of terms in the explanations.

Table 4-2: Total and distinct metalinguistic terms used for 13 grammatical features in 6 textbooks of German as a Foreign Language

Feature	*Total # of terms*	*Distinct terms*	*Ratio*	*Assumed (%)*	*Explained (%)*	*Linguistic domain*
Word order dep. clauses	36	20	1.8	1	99	syntax
Word order indep. clauses	30	20	1.5	83	17	syntax
Adverb placement	12	9	1.3	10	90	syntax
Separable prefix verbs	34	21	1.6	60	40	morphology/ syntax
Accusative	41	22	1.9	29	71	morphology
Dative	24	17	1.4	17	83	morphology
Negation of verbs (nicht)	6	6	1.0	33	67	syntax/sem.
Negation of nouns (kein)	18	16	1.1	13	87	morphology/ syntax/sem.
Modal verb choice	13	12	1.1	5	95	semantics
Modal verb placement	17	14	1.2	0	100	syntax
Auxiliary choice past tense	15	11	1.4	73	27	semantics
Participle choice past tense	38	20	1.9	21	79	morphology
Agreement	34	20	1.7	61	39	morphology
Range	35 41-6	16 22-6	0.9 1.9-1.0	83 83-0	83 100-17	
Mean	24.5	16	1.5	31	69	

As was true for the values sorted by book (Table 4-1), the range for the total number of terms for each feature is, with a value of 35, much higher than is the case for the distinct number of terms, which is only 16 – as was to be expected. The total number of terms was included to provide an insight into the (average) length of explanations in the grammar sections, or – in other words – how much space textbook authors or publishers allow for grammatical explanations and metalinguistic terminology. Unlike choosing “number of sentences” or “number of lines”, which would have been highly relative, the total number of terms is very precise.

When looking at the numbers of distinct terms, they may appear to be fairly high given the fact that they occur in the explanation for *one* grammatical feature. However, since metalanguage is here defined as “language about language” or “talk about language” (which, for example, includes terms such as “position”, “sentence” or “ending”) and not only as technical or scientific metalanguage, numbers of 20 and up are easily reached.

As was true for the textbooks, the combination of a higher range for the total number of terms and a lower one for the distinct number of terms indicates a higher variety for the *verbosity* found in the explanation for the individual features in contrast to the variety in terms of the distinct terms that were chosen.

For the total number of terms, the lowest values were identified for the explanations of the negations of verbs (nicht) (6), adverb placement (12) and modal verb choice (13); the highest number of total terms are used for the explanations of the concept of accusative (41), participle choice past tense (38), and word order dependent clauses (36).

For distinct terms, the lowest ranks are taken by (almost) the identical features: negation of verbs (nicht) (6), adverb placement (9) and, auxiliary choice past tense

(11). This does not hold true for the highest ranks, which are taken by different features, with the exception of accusative (rank 1, 22 terms): dative (21), word order dependent and independent clause, agreement, participle choice (20 each). The ratio of the number of total and distinct terms ranges from 1.0 to 1.9. The three highest values were found for participle choice past tense (1.9), accusative (1.9), word order dependent clauses (1.8), and agreement (1.7). The lowest values were found for negation nicht (1.0), modal verb choice (1.1), negation kein (1.1) and modal verb placement (1.2).

When considering the linguistic domains the features in question belong to, it can be noticed that, with one exception, features from the morphological domain can be found among the highest seven values (participle choice, accusative, agreement, separable prefix-verbs, dative). Only two out of the seven highest ranks are taken by non-morphological features, namely word order dependent and independent clauses (syntax).

For ranks 8-13, the features belong (mostly) to the field of semantics (auxiliary choice, modal verb choice, negation nicht), and syntax (adverb placement, modal verb placement).

Interestingly, features that are part of the same structure (perfect tense) or represent the same concept (case) showed rather high differences in terms of their value. For example, the morphological aspect of forming the past tense in German, namely participle choice, had a value of 1.9, while the semantic aspect, auxiliary choice, had a value of only 1.4. Likewise, the ratio of total and distinct terms was 1.9 for accusative and only 1.4 for dative. For word order, there was a difference of 0.3, with dependent clauses having a value of 1.8 and independent clauses a value of 1.5. However, for modal verbs that consist of a semantic (modal verb choice) and syntactic

(modal verb placement) component, these differences could not be observed with the former having a ratio of 1.1 and the latter a ratio of 1.2.

When looking at the concept of negation, the differences between the ratios for the negation with nicht and kein was only 0.1.

In terms of the dichotomy explained vs. assumed terminology, the highest three values for assumed terminology were identified for word order independent clauses with a value of 83%, auxiliary choice (73%) and agreement (61%); the lowest amount of assumed terminology was used for the features modal verb placement (0 %), word order dependent clauses (1 %), modal verb choice (5 %) and adverb placement (10 %). When sorted by linguistic domain, these features all belong to either syntax or semantics. While the same holds true for ranks 1 (word order independent clauses) and 2 (auxiliary choice), ranks 4-8 are taken by morphological features with values between 17% (dative) and 61% (agreement).

As was true for the ratio between distinct and total number of terms, different aspects of the same structure of the same concept were treated differently in terms of whether they were explained or not. For example, the components of the present perfect (auxiliary choice and participle choice) differed greatly in terms of how many terms were used and explained in the metalanguage: while for auxiliary choice, 73 % of the used terminology was assumed, only 21 % was assumed for participle choice. For the concept of case, about one third of the terms were assumed for accusative (29 %) and only 17 % for dative.

The most striking difference could be observed for word order in dependent and independent clauses: in the explanations for independent clauses, 83 % of the terms were assumed while only 1 % of the terminology was assumed for word order in dependent clauses. For modal verb choice versus placement, the percentages for

placement and choice were similarly low with values of 5 (modal verb choice) and 0 (placement).

Transparency and opacity

Table 4-3 offers an overview of the different degrees of transparency (or opacity), sorted by textbooks.

Table 4-3: Degrees of transparency / opacity of the metalinguistic terms for each textbook, in percent

Textbook	Opaque	Mostly Opaque	Semi-Opaque	Transp.	Semi-Transp.	Pseudo Transp.
Berliner Platz 1	62	6	6	13	13	0
Treffpunkt Deutsch	49	13	4	21	0	13
Wie geht's?	45	9	12	14	0	20
Sag Mal	47	17	5	14	0	17
Kontakte	46	4	7	29	0	14
Deutsch – Na klar!	54	9	1	18	0	18
Mean	50.5	9.7	34.2	18.1	2.2	13.7

For all textbooks, the value for opaque terms is clearly higher than is the case for transparent terms. Both the individual values for each textbook, which range from 45 % (*Wie geht's*?) to 62 % (*Berliner Platz*) for opaque terms as well as the mean value of 50.5 % are much higher than the values for transparent terms. These range from 13 % (*Berliner Platz*) to 29 % (*Kontakte*) and have a mean of 18.1.

Overall, one can say that these ranges are rather low, which means that the books are very similar in regards to the proportion of opacity of their metalanguage. This impression is confirmed when adding up all values for the categories opaque,

namely opaque, mostly-opaque, semi-opaque and transparent, namely transparent, semi-transparent and pseudo-transparent (Table 4-4):

Table 4-4: Transparency and opacity (all degrees combined) of metalinguistic terms for each textbook, in percent

Textbook	Opaque	Transparent
Berliner Platz	74	26
Treffpunkt Deutsch	66	34
Wie geht's?	66	34
Sag Mal	69	31
Kontakte	57	43
Deutsch – Na klar!	64	36
Mean	**66**	**34**

The mean shows that about two thirds of the terms that are used are part of the category opaque, while only one third is part of the category transparent. Again, the range between the highest (74 %) and lowest (57%) is rather low, which confirms the finding above that the textbooks are similar in terms of the proportion of opaque terms they are using.

Table 4-5 offers shows the proportion of opaque vs. transparent terms for each feature. For reasons of comparability, the values of all degrees of opacity and transparency are combined. The table further includes the proportions for assumed vs. explained terminology.

Table 4-5: Transparency and opacity (all degrees combined) of metalinguistic terms for each feature, in percent

Feature	Opaque (%)	Transp. (%)	Assumed	Explained	Linguistic domain
Word order dep. clauses	43	57	1	99	syntax
Word order indep. clauses	67	33	83	17	syntax
Adverb placement	51	49	10	90	syntax
Separable prefix verbs	50	50	60	40	morphology/ syntax
Accusative	76	24	29	71	morphology
Dative	87	13	17	83	morphology
Negation of verbs (nicht)	33	67	33	67	syntax/sem.
Negation of nouns (kein)	77	23	13	87	morphology/ syntax/sem.
Modal verb choice	77	23	5	95	semantics
Modal verb placement	43	57	0	100	syntax
Auxiliary choice past tense	87	13	73	23	semantics
Participle choice past tense	78	22	21	79	morphology
Agreement	58	42	61	39	morphology

As was true for the overall values for the textbooks, opaque terms are very dominant for almost all structures. With the exception of word order dependent clauses, opaque terms account for at least 50 % of all the terms that are used to explain the features in question. With 87 %, dative and auxiliary choice have the highest value for the number of opaque terms that are used, followed by participle choice with 78% and negation kein and participle choice with 77%, and accusative with 76%.

The lowest values can be observed for negation nicht (33%) word order dependent clauses and modal verb placement (43%), and separable prefix-verbs (50%).

In order to get a better understanding of the data, opacity will be analysed 1) in regards to features that are connected, either by being similar in nature or due to being part of the same structure (e.g. accusative vs. dative, perfect auxiliary vs. past participle, and so on) and 2) in connection with the fraction of assumed versus explained terminology.

One of the most drastic differences for similar structures in the same domain, namely syntax, can be observed for word order independent clauses vs. word order dependent clauses. For the former, 83 % of the terms are assumed whereas that is only the case for 1 % of the latter. At the same time, the proportion of opaque terms is much higher for word order independent clauses, with a value of 67 %, than it is for dependent clauses with a value of only 43 %. Consequently, it seems that opacity of metalinguistic terms does not seem to be a reason for explaining terms related to word order in independent and dependent clauses. High values for opacity and, at the same time, low values for explained items, can also be observed for 1) word order independent clauses (17% explained, 67% opaque), 2) auxiliary choice (23% explained, 87% opaque) and 3) agreement (39% explained, 58% opaque).

The reverse constellation of values, that is, a coincidence of high values for both explained terminology and transparency of metalanguage, can be observed for: word order dependent clauses (explained: 99 %, transparent: 57%), modal verb placement (explained: 100 %, transparent: 57 %), and negation nicht (explained: 67%, transparent 67 %). In sum, the data can be categorized as follows:

Table 4-6: Categories, sorted by transparency vs opacity as well as explained vs assumed terminology

High values opacity + high values explained terminology	High values opacity + low values explained terminology	High values transparency + high values explained terminology
Accusative (morphology)	Auxiliary choice (semantics)	Word order dep. clauses (syntax)
Dative (morphology)	Agreement (morphology)	Modal verb placement (syntax)
Participle choice (morphology)	Word order indep. clauses (syntax)	Negation nicht (mostly syntax)
Modal verb choice (semantics)		
Negation kein (various)		

In terms of linguistic domains, there are regularities for morphology and syntax: while morphology has high values for opacity and explained terminology, syntax has high values for explained terminology as well, but – unlike morphology – high values for transparent terminology.

For high opacity levels and low values for explained terminology, no pattern can be observed in terms of linguistic domain. However, the fact that three out of the thirteen structures fall into this rather counter-intuitive category needs to be kept in mind. Separable prefix-verbs and adverb placement are not mentioned in the table because their distribution of opaque and transparent terms was almost identical, which did not allow for comparisons.

Frequency and opacity of individual metalinguistic terms

In Table 4-7, the frequency of individual metalinguistic terms (as they appeared in the explanations in the textbooks) is provided, as well as their degree of opacity and the linguistic domain they (predominantly) belong to. Only those are listed that appeared at least three times across the textbooks; those that occurred only once or twice were not listed.

Table 4-7: Frequency and opacity of metalinguistic terms used, sorted from highest to lowest

Term	Freq. #	Opacity
verb (form)	32	opaque
sentence	22	pseudo-transparent
subject	12	opaque
direct object	9	mostly opaque
infinitive	9	opaque
position	8	transparent
ending	8	transparent
conjugated	7	opaque
plural (form)	7	transparent
pres. tense	7	semi-opaque
prefix	7	opaque
noun	7	opaque
adverb	6	opaque
clause	6	pseudo-transparent
pronoun	6	opaque
past part.	6	semi-opaque
stem	5	pseudo-transparent
regular	5	pseudo-transparent
preposition	5	opaque
case	5	opaque
ind. object	4	mostly opaque
conjunction	4	transparent
element	4	transparent
irregular	4	pseudo-transparent
acc. (case)	4	opaque
inflected	3	opaque
complement	3	transparent
singular	3	transparent
stem vowel	3	mostly opaque
perfect tense	3	opaque

This data confirms the findings from above in terms of the opacity of the terms. The five most frequently used terms are either opaque, mostly opaque or pseudo transparent, which is in line with the findings for both the textbooks and the individual structures. Both show clearly higher values for opaque terms than for transparent ones.

In addition to these quantitative findings, this table serves to provide a more comprehensive insight into the kind of terms that were used in the explanations of the individual features. When looking at these terms, it becomes obvious that the vast majority are of Latin origin, that is, most explanations of grammatical phenomena rely on Latin-based notions of grammar, even though there may be non-Latin-based alternatives. This also accounts for the high number of opaque terms that are being used in the explanatory sections of the textbooks. Again, this partially reflects the nature of the terms that were chosen for analysis since most of the names for the features themselves are of Latin origin, for example, dative, accusative or terms to describe the past tense (perfect auxiliary / participle).

In Appendix H, three example explanations that are representative of the Latin-based notion of grammar are presented and described in detail to further illustrate the nature of the metalanguage.

Discussion

Before the results are discussed in more detail, a brief summary serves to provide an overview of the most relevant findings.

Common features of the metalanguage across all textbooks:

- The number of distinct terms used for individual features is very similar across textbooks.
- The highest explanation strength could be observed in the linguistic domain of morphology, the lowest for semantics and syntax.
- The number of explained terms is significantly higher than the number of assumed terms; terms were often not explained when they first occurred but later on in the textbook.
- No connection could be observed between assumed/explained terminology and linguistic domain.
- Both across textbooks and for individual structures, there are more opaque than transparent terms:
    - High values for opacity and high values for explained terminology could be observed mostly in the domain of morphology.
    - High values for transparency and high values for explained terminology could be observed mostly in the domain of syntax.
- Opacity occurs mostly due to the fact that terminology is taken from Latin-based grammar.

- *Variable characteristics:*
- The number of total terms varies largely across textbooks and also for individual structures.
- High values for opacity and low values for explained terminology were observed for three structures overall, but they could not be connected to a specific linguistic domain.

When discussing these findings with the aim to identify what textbook authors seem to consider to be contributors to difficulty, the following can be said:

With one outlier (*Berliner Platz*), explanation strength, or the ratio between the overall and the distinct number of words that are used to explain a certain feature, was rather similar across textbooks with values around 1.5 (or a range from 1.4, *Treffpunkt Deutsch*, to 1.7, *Sag Mal*). This seems to hint at a similar understanding of how much *verbosity* is needed in the explanations of grammatical features and, thus, may be one shared way to facilitate the learning process or to reduce the amount of difficulty associated with metalanguage.

With regard to *Berliner Platz*, its outlier position with a value of 1.2 for explanation strength has to be assessed with pragmatic considerations in mind: depending on the overall length and density of the textbook, the space that authors (or publishers) can or want to devote to grammatical explanations can vary to a large extent. The fact that *Berliner Platz* has fewer pages than the other textbooks and, therefore, less space for grammatical explanations, confirms this assumption.

When looking at the average values for explanation strength and linguistic domain, morphology being the domain with the highest value is largely in line with research on difficulty (e.g. DeKeyser, 2005). However, whether this perception of

textbook authors is based on practical experience, knowledge of research on the matter or just the result of following textbook traditions either for the publisher or with an eye on competition on the market, cannot be answered without speculation.

Another common feature is the fact that all textbooks have a higher number for terms that are explained than for those that are assumed. This, again, seems to point at a similar understanding of metalanguage as a contributing factor to difficulty.

However, within textbooks, many terms were not explained when they were first used but in later chapters. Such inconsistencies may be due to different authors working on different chapters without coordinating the use of specific terms and whether or not they should be explained. These inconsistencies may also hint at the fact that deciding what should be explained and what can be assumed is not an informed process but rather random. Alternatively, this consistency could also be seen as a lack of attention in the editorial process because it should be in the focus of the people involved in this process to check the consistency in the use of specific terms.

Regardless of what the source of this inconsistency may be, when terms are explained in some chapters (and books) but assumed in others, it shows that there does not seem to be a common understanding of the (grammatical) background of the target group beyond the fact that the textbooks that were analysed were specifically designed for English-speaking learners of German, and thus, for specific countries. While economic interests of the publisher (i.e. the concern to reach as broad an audience as possible) can certainly not be ignored, the missing appropriateness or awareness of the target group can be seen as a lack of “adaptivity”, as described by Heringer and Keller-Bauer (1984) as one component that contributes to the difficulty of metalanguage.

This inconsistency in terms of if and when terms are introduced, can lead to even more problems: If a term is not explained at first but then later on, the learner may question his/her understanding of the term when first being introduced, which leads to insecurities. It may also lead to the question whether the term, when finally introduced, is explained comprehensively or if more facets to the term are to be expected. This comprehensibility of an explanation, which is "to lead to a clear and thorough understanding of a term", is another criterion that adds to the difficulty of metalanguage according to Heringer and Keller-Bauer (1984). In addition to this potential lack of comprehensibility in the learner's perception, it became very clear in the coding of the data that the explanations were often incomplete or unclear, which offers further support for the fact that this aspect may not be taken into account either, when decisions are made regarding the use of metalanguage in the textbooks. While more self-confident learners may not experience insecurities due to the *violation* of the rules formulated by Heringer and Keller-Bauer (1984), these learners may question the level of diligence with which the textbook was designed.

More inconsistency could be observed when assessing the opacity/ transparency of the metalanguage in connection with whether these terms were assumed or explained. While for morphology, a high number of opaque terms correlated with a high number of explained terms, which seems logical, this was not true for the domain of syntax. Here, high levels of transparency were connected with high levels of explained terms. These inconsistent findings seem to indicate that there is no connection between opacity / transparency and the perceived necessity to provide explanations for the terms. For three features – despite not belonging to the same linguistic domain – the values for opacity were rather high while the values for explained terminology were low. Consequently, textbook authors do not seem to generally perceive opaque terms as a source of difficulty, which contradicts both

common sense and research since opacity means that “there is no obvious relationship between the term and its referent” (Berry, 2010, p. 50).

Opacity is the result of the fact that terminology is mostly taken from Latin-based grammar, what is described as more scientific, distinct and precise (Berry, 2010). These features make it clear that not using Latin-based terms at all cannot be an option either. However, the explanations that were provided were not targeted at decreasing the opacity of the term by explaining or translating what the Latin words actually mean – despite their communicative nature. Instead, this communicative *value* of the Latin terms was completely left out and, with it, its potential to shed light on the actual meaning of the words (and their grammatical role). *Shedding light* is here based on the definition of opacity and transparency that was used in this paper; future research could focus on the learners’ perceptions of what does and does not contribute to making a term more or less opaque.

This consistent use of Latin-based terminology across textbooks further implies a rather traditional approach to language and language learning (Funk, 1995), which is a strong contrast to the fact that all books claim to focus on communication and some of them even explicitly mention the approach of Communicative Language Teaching. When looking at the grammar sections only, it would be impossible to identify this communicative approach; instead, one would assume that all textbooks seem to put strong emphasis on structural or formal grammar.

On a more general level, such consistency is surprising because the strong connection between metalanguage and how languages are learned are no longer as uniform as they used to be up the 1980s. The dogma, that was described to have disappeared by then (Funk, 1995), still seems to be existent in these textbooks. This is not to say that the classes or courses in which these books are used are equally

traditional in nature. However, the more *modern* or communicative the approach, the stronger the discrepancy between what is done in class and how grammar is explained will be perceived by the learner. This discrepancy becomes even more problematic if the communicative classroom does not put strong emphasis on explicit instruction and assigns reading grammar sections as homework or for self-study.

Pedagogical implications

As was pointed out, the use of Latin-based terminology is – to a certain degree – unavoidable to ensure the specificity of the terms. Furthermore, not every Latin-based term has a non-Latin-based alternative. However, certain measures could be taken to decrease potential difficulties associated with such excessive use of Latin-based terminology:

- Adjusting the use of metalanguage to the target group

Not only is a strong Latin-based notion of grammar a manifestation of the authors' understanding of how languages are learned, but it can also be seen as a manifestation of the perceived learner. With – on average – one third of the terms being assumed in the textbooks, a learner is constructed who (to a certain extent) is perceived as having little difficulty navigating such (traditional and explicit) notion of grammar, including the terminology, either due to having received such training at the secondary level, or by possessing explicit knowledge in other languages or in linguistics. These assumptions are rather Euro-centric since explicit grammar instruction in the learners' L1 can be observed more frequently in European syllabi for secondary schools (especially Germany) than is the case in North-American high school curricula. When having a

closer look at the biographies of the textbook authors, such Euro-centrism is not surprising as most of them were socialized in Germany.

By being more aware of the target group and what their background may be (and thereby ensuring “adaptivity” as formulated by Heringer and Keller-Bauer, 1984), further changes to Latin-based terminology could be made, for example, replacing Latin-based terms with non-Latin ones, where possible, as well as increasing the quality of the explanations. While this was not in the focus of this study, the analysis of the explanations showed that they were often incomplete, unclear, or ambiguous and sometimes lead to the use of more metalinguistic terms although they could be avoided. Although this claim requires systematic future research, it has to be kept in mind when assessing what type of learner is assumed or what is expected from them.

- Making use of the communicative value of Latin-based terminology

From the analysis it became clear that Latin terminology is often presented as arbitrary or opaque, that is, without a connection between the signifier and the signified. However, many Latin-based terms are very communicative in nature because they tell a lot about the function and effects of individual grammatical features. If that connection was made more transparent, the supposed arbitrariness and technicality of grammatical terminology may be reduced and thus the relationship between content- or communicatively oriented unit-contents and grammar sections may be perceived as *smoother*.

While, generally speaking, language is used to allow communication between two speakers, grammatical terminology should be presented as a way to establish communication between the language and the learner.

- Decreasing inconsistencies in the use of metalinguistic terms within and across textbooks

As was pointed out, metalinguistic terms are not always explained when they are first introduced, which violates the rules for reducing the difficulty associated with metalanguage as identified by Heringer and Keller-Bauer (1984). Therefore, it could be a first step for textbook authors or publishers to establish the rule that terms are explained when they first occur. However, not only would this require coordination amongst contributing textbook authors, it also requires a common understanding of what should be explained and why as well as how. Instead, inconsistencies could be tackled by practitioners directly, for example by directing teachers' attention to these inconsistencies in the use of metalanguage. Once teachers are aware of the issues associated with the use of metalanguage (such as inconsistencies or low quality of explanations), the implementation of metalanguage in the classroom could be adjusted. For example, teachers could make sure to support a common understanding of metalinguistic terms when they first appear in the textbook or to complement or replace existing explanations with their own.

Alternatively, textbooks could make more extensive use of glossaries or create handbooks that are solely devoted to explaining metalinguistic terms. While one may argue that general grammar books that students can purchase may serve that purpose, these books are often intimidating because of their (usually high) price and the in-depth explanations of not only terminology but grammar in general. Due to the similarity of introductory-level textbooks with regard to the grammatical structures they introduce and which distinct terms they use in their explanations, as was demonstrated, such handbooks may even be designed to be used not for a specific textbook only but for beginner textbooks in general. While this measure is certainly the one that requires a

high (and possibly unrealistic) amount of resources, it may also be one of the few measures that could contribute to not only decreasing (or compensating for) inconsistencies in the use of metalanguage within a textbook but also across textbooks.

- Teaching teachers how to deal with metalanguage in textbooks

This final suggestion repeats the call for the inclusion of strategies for teaching and dealing with metalanguage in teacher training curricula (Ivancic, 2010). Given the diverse challenges associated with metalanguage, as described above, teacher intervention may be the most direct and successful way of dealing with these challenges (as was already mentioned in the previous suggestion). Courses (or other formats) that present ways to deal with instructional material should not only focus on how to use textbooks on a general level but specifically on the use of metalanguage. Instead of, or in addition to, (passively) presenting teachers with the issues that can occur when metalanguage is used, teacher trainees could be encouraged to critically analyse and assess the metalanguage themselves. Ideally, they could be confronted with metalinguistic explanations for a language they themselves do not speak to directly sensitize them for problems that may occur so they can develop ways to facilitate the use of metalanguage in both the classroom and in textbooks.

However, for all these measures, it is necessary for textbook authors as well as other involved parties, such as publishers, to be aware of the problems associated with the use of metalanguage and how it may be connected with the concept of difficulty. Hopefully, studies like this one will contribute to achieving this goal.

Conclusion

This study aimed at providing a comprehensive analysis of the metalanguage that is used in textbooks for German as a foreign Language and how specific attributes of metalanguage may contribute to difficulty. It has taken a very fine-grained approach and identified both patterns and inconsistencies in the use of metalanguage and identified areas that could be used to increase learner-orientation in textbook design. However, it also has limitations and generated ideas for future research.

The small number of textbooks that were analysed is certainly a limiting factor of this study since it is difficult to draw representative conclusions from the analysis. Likewise, a higher number of structures would increase the level of representativeness, especially in regards to patterns or connections between which terms tend to be explained and which are not.

Furthermore, the use of Latin-based terminology is assessed without including the voice of those who are directly affected: the language learner who is confronted with the books that were in the focus of the analysis. Future research could investigate how learners perceive the use of Latin-based terminology and what influence it has on their acquisition process and learning.

As was hinted at several times, the quality of the explanations was often problematic. Since this was not part of the research questions of this study, future studies could systematically examine textbook explanations for grammatical features; for example, by focusing on the relationship between pedagogical and scientific rules. It would serve as another way of contextualizing findings from studies like this one, where a quantitative approach was taken to the use of metalinguistic terms.

For further contextualization, future research should focus on 1) the actual use of textbooks, especially the implementation of the grammar sections by German as a

foreign language instructors and 2) textbook design: How do authors go about the use of metalanguage when designing a textbook? Do they have rules for how much or how little metalanguage is used and for what kind of metalanguage they use for which features? Are there rules for the use of metalanguage that need to be followed by the authors that contribute to the different chapters of the textbook? What role does research play in textbook design and how is the connection between research and practice established?

With studies like these, a more comprehensive understanding of both the use of metalanguage and the concept of difficulty could be achieved, which can then lead to more research-informed and learner-oriented decisions in textbook design.

Chapter 5

Study 4: The Sequencing of Grammatical Features in German as a Foreign Language Textbooks

This chapter focuses on the pedagogical approach to difficulty. It presents a study on how grammatical features are sequenced in beginner textbooks for German as a foreign language. It aims to find out whether a typical sequence of grammatical structures exists across different textbooks and which aspects may have an influence on this sequence. Normalization was used as a new method to account for the relativity of the data and make the findings comparable.

Introduction

Depending on the institutional context, language instructors may be facing different levels of involvement when it comes to syllabus design. Most teachers, be it at the secondary or post-secondary level, need to follow learning objectives that are part of a broader curriculum, which is either designed by program coordinators (university) or by governmental curricula (public schools). In other words, language teachers can be considered “consumers” (Bell, 1983) of syllabi, or curricula, that other people have designed rather than being actively involved in the design process. Despite the various ways of both curriculum (e. g. London School, Lancaster School and Toronto School, Rahimpour, 2010) and syllabus design (for an overview see Krahnke, 1987), beginner language classes often use textbooks as the basis for making syllabus- or even curriculum-related decisions (Rodrigues, 2015; Byrd, 2001)

With a textbook being in the focus, decisions on what to include, how to include it and when to include it are often made by textbook designers, who themselves often are not practitioners in the classroom, and not by teachers. Of course, teachers can

still interact with the textbook and make their own decisions based on how they implement what is offered as long as they take into account the overall syllabi in coordinated (e.g. university) or government programs (secondary schools); however, with regard to beginner classes, there is one aspect of textbook design which does not leave much room for adjustments: the sequencing of grammatical forms and structures (often referred to as *gradation*). While, at higher levels, most textbooks may be planned based on specific gradation principles, they can be used modularly (i. e. one can choose which unit is used at what point). The sequencing in beginner textbooks, however, usually does not allow for such freedom since gradation is, *per definitionem*, progressive in nature with its main purpose being the decrease of learning difficulty and an increase of learning outcomes. At the beginner's level, gradation is crucial since it ensures that structures are introduced in a way that takes into account the implications, that is, the knowledge that is necessary to have when more complex or advanced structures are introduced. For example, students need to be familiar with the concepts of independent clauses before dependent clauses are introduced. Otherwise, they would not be able to form complete sentences.

In the 17th century, Comenius was one of the first who recognized this role of gradation; he stated that "systematic gradation reduced the difficulties of language learning by distributing the extensive material of a language into steps arranged in specifically prepared texts in which everything progressed, not by leaps and bounds, but gradually" (in: Mackey, 1965, pp. 204-205).

With this definition, the connection to the concept of difficulty, which is of direct relevance for this thesis, is explicitly addressed as the core motivator for decisions related to gradation or sequencing. While, back then, sequence-related decisions were mostly based on common sense, sequencing was soon included in instructional

considerations on a theoretical level (e.g. Krashen & Seliger, 1975) and in empirical research. The latter mainly focused on identifying developmental sequences in second language acquisition contexts (e.g. Clahsen et al., 1983; Dulay et al., 1983; Dulay & Burt, 1973). Strongly connected with these findings is the acquisitional approach to the concept of difficulty: easy structures are acquired early and difficult structures are acquired late (Collins et al., 2009).

With gradation playing such a crucial role in textbook design and, therefore, on the instructional level, it is worth examining at how gradation, or grammar sequencing, is implemented in current beginner textbooks of German as a foreign language. In particular, the question will be answered whether a typical sequence exists across textbooks; a question that has so far remained unanswered by research.

In the following literature review, an overview of the history of gradation, how it was perceived and how it can be implemented in textbooks are reviewed. Research on L2 developmental stages is presented in the following section, which concludes with practical suggestions that can be drawn from research for textbook design. The final section connects gradation and the concept of difficulty before the focus will be on gradation research that has been conducted in the field of German as a foreign language.

Literature review

The history of gradation and how gradation can be implemented in textbooks

Traditionally, sequencing decisions were based on common sense, past experiences, or assumptions about the relative difficulty of grammatical features. Despite the fact that the purpose of gradation, namely decreasing learning difficulty and increasing learning outcomes, as stated above, was formulated by Comenius as early as in the 17th century, with his death, principles like these were not found in language study material up until the 19th century (Mackey, 1965). While the idea of gradation being necessary for effective language learning remained, the principles that were applied varied greatly from a "natural […] and logical […] [connection]" of questions (Heness, 1885), to Gouin's (1892) "principles of association psychology" (both in: Mackey, 1965, p. 205), which relied on topic-based associations such as Home, Nature, Society or Science. A more linguistic-specific focus developed in the early twentieth century, when syntax and morphology became more of a focus in addition to semantics. Palmer (1917), for example, explained that by classifying language into "groups" or domains such as morphology etc., "we are enabled to teach and to learn facts on a wholesale scale […]. It is only when we realize the nature of the agreements and differences shown by the scheme of classification that we may be said to understand the subject of our study" (p.42).

These decisions were usually based on the linguistic complexity of the forms such as, for example, the past perfect progressive in English *(I had been waiting*) was considered more difficult than the present progressive (*I am waiting*). Complexity, in this context, is often defined as the number of decisions that a learner needs to make in order to arrive at the correct outcome (Housen & Simoens, 2016).

When looking at grading, as manifested in textbook syllabi or orders of introduction of certain features, one can – in general – differentiate between *synthetic* and *analytic* (Wilkins, 1976) approaches to syllabus construction. According to Wilkins, the former is based on structural considerations in the target language, which is further broken down into three sub-principles: 1) new structures need to be introduced in a way that they refer back to old structures, 2) simple structures are to be introduced before complex structures and 3) the connection of interaction between linguistic structures needs to be taken into account. It is based on structural, or morpho-syntactic, grading.

Analytic syllabus grading is "organized in terms of the purposes for which people are learning and the kinds of language performance that are necessary to meet those purposes" (Wilkins, 1976, p.13). This type can be considered to focus more on the communicative aspects of language learning, and may find its manifestation in the form of task-based organization.

Research on L2 developmental stages and the connection between gradation and the concept of difficulty

With the emergence of empirical evidence relating to L2 developmental sequences such as morpheme order studies (Dulay & Burt, 1973), negation (Dulay et al., 1982), or word order (Clahsen et al., 1983), SLA researchers in the 1980s began to consider the possibility of pedagogical sequences being based on learnability issues. Even the synthetic syllabi mentioned above did not take into consideration the learner and learnability, but almost exclusively looked at structure-inherent features that were used for grading.

Pienemann (1985) for the first time comprehensively approached gradation, or syllabus construction, in the context of learnability with the objective “to provide the syllabus designer with more precise principles for grading teaching material which are related to simplicity as defined within the framework of language learning” (p. 24). These principles are based on more general, psychological considerations, which can be applied to the context of (second) language learning:

1. Do not demand a learning process which is impossible at a given stage (i.e. order of teaching objectives be in line with stages of acquisition)
2. But do not introduce *deviant* forms
3. The general input may contain structures which were not introduced for production

(Pienemann, 1985, p. 63)

Strongly connected with the learnability of features is their teachability; this discussion mainly revolves around the effects of instruction and if grading in instruction is necessary at all, especially given the fact that there seems to be a natural order in which certain features of a language are learned or acquired (for a more in-depth discussion see Pienemann, 1984).

With respect to the topic of difficulty, pedagogical sequencing in textbooks is strongly connected to what Collins et al. (2009) refer to as the "acquisitional perspective". For example, the morpheme order studies (Dulay & Burt, 1973) or Processability Theory (Pienemann, 1989) define difficulty in terms of “whether a structure is ‘early’ or ‘late’ acquired” (Collins et al., 2009, p. 336). According to these studies, the acquisitional process follows a fixed order regardless of when a specific feature is introduced in institutional settings (s. also Ellis, 1989; Pica, 1983; Perkins & Larsen Freeman, 1975). A plethora of studies were motivated by these findings and aimed at identifying natural orders for linguistic elements/constructions other than morphemes, for example, constructions such as negation (Dulay et al., 1982), or

relative clauses (e.g. Ammar & Lightbown, 2005; Izumi & Izumi, 2004). A commonly expressed critique about these studies, namely the fact that they were lacking theoretical explanations for the orders they identified (Gregg, 1984), leads to research that is of particular relevance to this study – the work of the ZISA group (Clahsen et al., 1983). Using German word order acquisition to prove that there are 'natural sequences', "this group explained sequences through underlying processing strategies that act as constraints on transformations" (Baten, 2011, p. 456). A few years later, Pienemann (1989) refined these findings and theoretical considerations in the framework of Processability Theory (PT), in which he postulates that learners go through a series of predictable developmental stages when learning a foreign language. These stages are: 1) no procedure (lemma access), 2) category procedure, 3) phrasal procedure, 4) sentence procedure, and 5) subordinate clause procedure (if applicable) (Pienemann, 1998, p.80). In the development of a learner, stages cannot be skipped, that is, it is impossible for a learner to show phrasal procedures before category procedures are being mastered. Difficulty can therefore be understood in terms of developmental readiness and the features involved, or, as Ellis (2008) puts it, "learning difficulty and the sequence of acquisition are determined by the nature of the processing procedure required to produce a specific grammatical feature" (p.11).

Although some studies in the field of German as a foreign or second language have shown inconsistencies with Processibility Theory (e.g. Dimroth & Haberzettl, 2007; Haberzettl, 2005 for German as a second language or Diehl et al., 2000, 2002 for German as a foreign language), it is commonly agreed upon that a natural acquisitional order does exist. For German, these orders have been identified for verb placement, sentence negation, tense and case (e.g. Baten & Williams, 2012; Pienemann & Kessler, 2011; Baten, 2011; Jansen, 2008; Ballestraccim 2005; Tschirner, 1999; Pienemann & Hakansson, 1999). Variation often seems to be

associated with the learners' L1 (e.g. Murakami, 2013 for morphological acquisition; Ellis et al., 2012 for German gender, or Haberzettl, 2006; Vainikka & Young-Scholten, 1996 for acquisitional differences of verb placement).

It may now be tempting to assume that grammatical sequencing in textbooks is not important since these patterns do not seem to be broken by aspects other than the learners' L1. While it holds true that the overall acquisitional sequence does not depend on instructional grammatical sequencing, various studies have found that being confronted with an order that is different from the natural one will lead to learning difficulties (Haberzettl, 2006; Diehl et al., 2000; Tschirner, 1999). Even though the strength of this connection may vary depending on the different learner groups, that is, findings on acquisitional sequencing in naturalistic settings may be less relevant for explicit grammar instruction for educated adult learners than for younger learners in immersive, implicit settings, for example, the connection cannot just be dismissed. Instead, these findings are (or should be) of relevance for textbook design; specifically, the order in which grammatical sequences are introduced should be determined by findings on natural orders for the language in question (for an overview of how different aspects of theory, research and practice are connected in a model for second language teaching see Stern, 1983).

Research on German L2 textbooks

Research on grammatical sequencing in textbooks for German as a foreign or second language is usually contextualized within the acquisitional perspective. Very few exceptions examine sequencing without referring to PT; those that do usually draw on the Common European Framework of Reference for Languages (e.g. Winkler, 2011; Westhoff, 2007), although this use of the CEFF has been criticized by

Kwaakernaak (2007), who does not consider the CEFR to be of prescriptive nature from a theory-driven perspective (p. 84). Instead, he suggests to rely on acquisitional, L1-sensititive research as described above when making decisions on grammatical sequences in textbooks.

In teaching materials for Dutch and Finnish learners of German, Tammenga-Helmantel (2012) focused on whether the "observed grammar sequences are in accord with general acquisition patterns for verb placement (I), sentence negation (II), tense (III), and case assignment (IV)" (research question 1). She also aimed at identifying the individual orders in which grammatical items are introduced in Dutch, Finnish and global German as foreign language textbooks.

For the analysis, five textbooks for Dutch learners of German, four textbooks for Finnish learners of German and three global textbooks were chosen. Assessing the congruence between acquisitional sequences and the order in which items were introduced in the textbook, the author found that overall "the analyzed teaching materials present tense, verb placement, and case assignment [are] in accord with general empirically based acquisition sequences" (p.72) Exceptions could be observed with regard to case-marking acquisition in the Dutch materials, sentence negation in both Finnish and global materials as well as the placement of verbs in Finnish materials.

In regards to linguistic domains or categories within that domain, distributional patterns could be identified:

- Tense order: present – perfect – past
- Verbal constructions: *haben* / *sein*, weak verbs – strong verbs – modals – verbs in the perfect tense;

- Morphological case: nominative – accusative – dative - genitive

The author concludes that by focusing on morphologically simple structures, and thereby reducing the intake-load, and on frequently used structures that allow for an inductive approach, a "learner-centred and communicative approach to foreign language teaching [is fostered]" (p. 75).

Aguado (2012) analysed the grammatical progression in five GLF textbooks and found that some parts of the sequencing were in line with acquisitional patterns, while others were not. She sees both informed, research-oriented sequencing as well as coincidence, that is, sequencing that does not seem to follow any acquisitional patterns but is intuitive in nature (p.15), as possible explanations for the congruence she identified. With regard to syntax, she identified subject-verb patterns as well as questions (verb subject) to be at the very beginning of the sequences she analysed in all textbooks. More variety was observed for separable verbs which were not introduced at a typical position in the textbooks. A similarity between textbooks, however, was the fact that separable prefix-verbs are introduced first, then (or at the same time) modal verbs and finally the perfect tense. Not only is this order the same in all textbooks, it is also in line with acquisitional orders identified by Diehl et al. (2000). When it comes to tenses, no such congruence could be observed. While the simple past tense (Präteritum) is acquired later than the Perfect tense, this order could not be identified in the textbooks. She, again, concludes the lack of congruence to be the result of the gap between research and practice (p. 9).

More consistency was observed by Maijala (2010) who compared grammatical sequences in Finnish and so-called global textbooks for German as a foreign/second language that are not targeted at a particular group of L1 speakers. The introduction of the main verbs *haben* and *sein*, verb conjugation, personal pronouns, definite and

indefinite articles occur very early across textbooks; differences were observed for the introduction of possessive pronouns. While they were among the first features to be introduced in the global textbook, they were introduced later in textbooks designed for Finnish learners of German.

Once again, these findings show that textbook design is not consistently informed by SLA research. While some structures are apparently intuitively perceived as more or less difficult, and thereby introduced earlier or later, instances where this is not the case (e. g. possessive pronouns) tend to be introduced at very different positions in the textbooks.

Despite the direct connection between grammatical sequencing as part of the acquisitional approach and the concept of difficulty, no research could be found which contextualized the findings accordingly, with one exception for German: Neuner (2003, 1996) establishes a connection between difficulty and grammatical sequencing in textbooks for German; however, these suggestions refer to the audiolingual method and are theoretical (or prescriptive) in nature and not empirically driven. He suggests determining a grammatical sequence based on the complexity of the feature, that is, less complex structures are to be introduced early and more complex ones later on.

None of the textbooks that have been analysed in previous studies were designed for English learners of German at the university level. So far, only individual features were of relevance; that is, they did not identify a typical *overall* order but only the position of individual features or small groups within these features. Based on the identified gaps in research, this study aimed to answer the questions:

1. Is there a typical sequence of grammatical structures in textbooks for German as a foreign language?
2. What factors influence sequencing?

Typical sequence is understood as an identifiable order, or patterns, in which grammatical structures are introduced *across* different textbooks, that is, do textbooks introduce the same structures at the same point / time? The structures that will be examined are the ones that were in the focus of Studies 2 and 3 to allow for comparability of the findings: subject-verb agreement, separable prefix-verbs, negation (*nicht* and *kein*), accusative with direct objects, dative with indirect objects, verb placement in dependent and independent clauses, modal verb choice and placement, adverb placement and choice of auxiliary and choice of participle in the perfect tense.

Since most textbooks describe themselves as either communicative in nature or as focusing on different communicative contexts, one factor that will be analysed is the question whether the identified order seems to be analytic or synthetic according to the criteria formulated by Wilkins (1976).

In addition, as was pointed out above, difficulty is connected with sequencing in two ways: 1) according to the acquisitional perspective (Collins et al., 2009), difficulty is determined in accordance with the point of introduction: easy structures are acquired early and difficult ones are acquired late. Hand in hand with this approach goes 2) research that has identified fixed acquisitional orders (s. literature review above). Therefore, acquisitional patterns will be assessed as another potential factor to influence sequencing.

Method

Textbooks

The contents of nine textbooks for beginner learners of German were selected for analysis. The textbooks were partly chosen based on an unpublished survey conducted by a Canadian professor of German as a foreign language (K. Misfeldt). A

total of 21 Canadian universities provided (among other related pieces of information) the title and edition of the textbook they use for introductory German classes. A total of nine different textbooks were reported to be used. At the time of data collection, seven out of these nine were accessible (without major shipping delays). These were either monolingual in German (i. e. global), or bilingual in English and German. The monolingual textbooks are *Menschen (Habersack et al., 2012)* and *Schritte (Penning-Hiemstra et al., 2017)* and the bilingual ones are *Berliner Platz 1 (Lemke et al., 2009), Treffpunkt Deutsch (Gongolweski et al., 2013), Wie geht's? (Sevin & Sevin, 2007), Sag Mal (Anton et al., 2014),* and *Kontakte (Tschirner et al., 2012)*. The two textbooks that were not accessible (one monolingual one and one bilingual one) were replaced by *Deutsch – Na klar! (Di Donate & Clyde, 2015)* (bilingual) and *Studio D (Funk et al., 2014)* (monolingual) because they had been used by myself for university level courses in Canada at the introductory level (but were not reported in the survey because they were used when no survey was conducted).

In contrast to Study 3, the monolingual textbooks that were mentioned in the survey were not excluded from analysis because the language used for grammatical explanations and annotations was not considered to be a variable with regard to identifying a potential sequence.

All of these books are described as "introductory" in nature or "for beginners", which means they all pursue the A-level according to the Common European Framework of Reference for Languages. Information on the amount of time each textbook is supposed to be covered in was not provided but due to the nature of data collection and analysis (s. below), this was not considered to be relevant.

All textbooks are communicatively oriented; some of them explicitly refer to Communicative Language Teaching in their introduction (e.g. *Kontakte*), but all of them

mention communication or communicative abilities in the target language as their major goal. The unit overview in all of these books is structured based on communicative contexts that show a rather low level of variability. Conversational settings all revolve around: introducing oneself, talking about routines and hobbies, living arrangements and the city one lives in, shopping, scheduling events (with friends or professionally), travel, transportation, family, the weather, food and dining, university and professional life.

In addition, all textbooks exclusively aim to teach German as a foreign and not as a second language. None of the textbooks provide any information on how sequencing decisions were made.

Analysis

The analysis began by examining the introductory overview at the beginning of each textbook. The corresponding grammar sections served to confirm or clarify the assumed meaning of the listed items in the table of contents since the relevant structures were not always mentioned individually and explicitly but were either part of a more comprehensive understanding or were labelled differently. For example, subject-verb agreement was usually part of sections called “verb conjugation in the present tense” or personal pronouns. Choice of auxiliary and choice of participle were usually treated under headings such as “Speaking about events in the past".

Another issue that needed to be taken into account was the formulaic occurrence of the features; in some textbooks, the introduction of “Wie heißen Sie?” (= What is your name, formal address) appears under the heading “Word order in questions”. However, the corresponding grammar section does not explicitly address word order but refers to the expression as a type of formula and provides more input

on the appropriate pragmatic setting to ask this question and how to respond to it. Despite this pragmatic or semantic reference, headings like these are described as grammar or structural knowledge in the textbooks. Consequently, these structures were excluded from the relative order and, thus, the analysis.

When a feature occurred multiple times (in different contexts), only the first relevant occurrence was analyzed. An occurrence was considered to be *relevant* when it was mentioned in reference to a grammatical structure and not to pragmatic or semantic ones (s. above).

Reliability was ensured by having two teachers (of German and English who both spoke both languages) identify first occurrences and analyse them. In cases where there was no overlap, these instances (seven in total) were discussed until an agreement was reached. In addition, textbooks rarely cover the same total number of grammatical features, which makes comparability difficult. In the Methods section it will be discussed how this relativity was accounted for.

In order to compare the findings of this study with findings from previous studies on different approaches to difficulty, the identification of a (relative) sequence was limited to the grammatical structures that were in the focus of studies 1 to 3.

Table 5-1 presents an overview of these features. The column farthest to the right provides examples of typical ways in which the target features are referred to in the textbooks. The descriptions in Table 5-1 are provided by the author.

Table 5-1: GFL textbook labels for target grammatical features

Feature	Description	Sample Designation in Textbook
Subject-verb agreement	Due to its complex morphology, verb endings in German are different for almost each personal pronoun in singular and plural	Verb conjugation in the present tense
Separable prefix-verbs	Some verbs need to be separated when conjugated	Separable prefix-verbs / separable verbs
Negation nicht	Verbs are negated by using *nicht* (not)	Position of *nicht* in a sentence / Negation
Negation kein	Nouns are negated by using *kein* (=*no*)	(possessive) Articles (ein/kein)
Accusative case with direct objects	Direct objects require the use of the accusative case; words change their endings based on case	The accusative case / Personal pronouns in the accusative case
Dative case with indirect objects	Indirect objects require the use of the dative case; words change their ending based on case	The dative case / Personal pronouns in the dative case
Verb placement in independent clauses	In independent clauses (or sentences), the verb needs to be in 2nd position	Independent clauses / Coordinating conjunctions (often combined with dependent clauses)
Verb placement in dependent clauses	In dependent clauses, the verb is pushed to the final position in the sentence	Dependent clauses / Subordinating conjunctions (often combined with independent clauses)
Modal verb placement	When modal verbs are used, they take the 2nd position in the sentence, while the main verb is pushed to final position	Modal verbs / Position of model verbs in the sentence / "Satzklammer" (= verbal bracket)
Modal verb choice	Modal verb choice often depends on nuances in meaning, which is similar to English	Modal verbs / Choosing the correct modal verb
Adverb placement	When a sentence contains a time and a place adverb, the adverb indicating time needs to precede the adverb indicating place	Time expressions / Locations / Word order: time before place/ Adverbs
Choice of auxiliary in the perfect tense	In German, the perfect tense has two components: an auxiliary verb and a participle; the auxiliary verb is either *haben* (have) or *sein* (be) depending on the verb type (state versus movement)	The perfect tense / Talking about events in the past
Choice of participle in the perfect tense.	In German, the perfect tense has two components: an auxiliary verb and a participle; the participle may contain the prefix "ge" or not, end with a "t" or a "en"; stem vowel changes are also possible and subject to whether the verbs are weak or strong	The perfect tense / Talking about events in the past

The following steps were taken in order to identify a typical sequence across the nine textbooks:

1. In each textbook, the structure overview was analysed to identify which of these features actually qualified as grammatical in nature, that is, whether the explanations referred to structural characteristics (as opposed to semantic or pragmatic use).

2. All identified topics (i.e. structures) in the textbooks were then numbered, which can be considered to be the absolute order of all topics in each textbook.

3. To make comparisons across textbooks, the absolute order was normalized using min-max scaling. Normalization is a way to adjust values to ensure that they are comparable by reaching a scale that is notionally common. Normalization was necessary for the following reason: Although all textbooks were advertised as being introductory/pursuing the A-level, there were large differences in terms of the number of grammatical structures covered by the textbooks. Due to the resulting differences in range, normalization was necessary before the average position (i. e. median) for each feature could be computed. Without normalization, the position of the features could not have been compared and, thus, a representative mean or median could not have been reached.

   Min-max scaling is based on the following formula:

   $$\text{Normalized value } x' = x - \min(x) / \max(x) - \min(x)$$

   This form of rescaling leads to values between 0 and 1. In the context of this study, these values can 1) be used to compare the different values for each textbook and 2) the values express the time of introduction in the textbook. A

value at or close to 0 means that a feature was introduced very early in the textbook, while a value of or around 1 means that it was introduced at the end or towards the end of the textbook. This reference to the position within the *textbook* makes it clear that the total number of grammatical structures in each book is considered in the steps involved in normalization. Before an average relative order can be determined, the absolute order needs to be identified, which includes every grammatical structure in every textbook.

4. Taking all data sets, that is, all textbooks into account, measures of central tendency (range, mean and median) were computed for each feature using the normalized absolute order with values from 0 to 1.

5. Based on the findings, the median values (and not the mean) were then sorted from lowest to highest, which was considered to be the typical sequence of grammatical features in textbooks for German as a foreign language. Using the median instead of the mean values was based on the fact that the former is robust against outliers whereas the latter is not.

6. In a final step, the findings will be compared with the aspects that have been mentioned as potentially influential on sequencing in textbooks, namely analytics vs. synthetic design, difficulty level, and acquisitional orders identified in the literature.

In addition to mean and median, the range will serve to further contextualize and interpret the data in order to reach the highest approximation to an overall order or distribution of the features across the different textbooks. When looking at how data is distributed, the range contributes to a more comprehensive understanding of how spread out the values are in the specific sets of data, which means it shows what the earliest and what the latest occurrence of a feature is for each textbook.

Although it is normal practice to compute standard deviations as a measure of the variability of the data, this was not appropriate, given a sample size of 9 books and only 13 structures. However, some information about variability in this sample can be gleaned from the range (i.e., maximum score minus the minimum score): while a small range hints at low variability in the data, which means there are very few differences in terms of when features are introduced, a high range hints at the opposite, namely high variability. In the latter case, there would be less consistency in terms of point of introduction across textbooks.

Previous research has not used a method that allows for a comparable representation of the position of specific features within a textbook. Most studies relied on representations that were relative within the textbook in question, for example by referring to information such as “unit 4 out of 11” or by determining different categories such as early and late, which relied solely on the researcher’s classification decisions (cf. Tammenga-Helmantel, 2012; Maijala, 2010). Not only are these approaches very subjective, they also do not allow for precise comparisons across textbooks since values such as 4 out of 11 for one textbook and 6 out of 16 for another are hard to compare.

Results

Identification of an overall typical sequence

In this section, normalized values for the features in question as well as the mean, median and range will be presented to answer the question whether or not a typical overall sequence could be identified. Tables 5-2 and 5-3 provide an overview

of the median values of each feature. Table 5-2 sorts the median values of each feature by textbook and table 5-3 from lowest to highest values across textbooks:

Table 5-2: Normalized values for each feature, sorted by textbook

	Wie gehts?	Berliner Platz	Kontakte	Sag Mal	Deutsch Na klar!	Treff punkt	Menschn	Studio d	Schritte	Range
Agreement	0	0.065	0.186	0.043	0.057	0.051	0	0.118	0.097	0.186
Order independ. clauses	0.049	0	0.220	0.087	0	0.093	0.210	0	0.032	0.220
Order dependent clauses	0.268	n.a.	0.424	0.652	0.314	0.423	0.742	0.841	0.801	0.573
Accusative	0.121	0.516	0.271	0.065	0.086	0.258	0.345	0.353	0.557	0.741
Dative	0.195	0.645	0.525	0.435	0.143	0.639	0.688	0.735	0.772	0.629
Separable prefix-verbs	0.488	0.355	0.237	0.413	0.257	0.371	0.531	0.471	0.677	0.440
Modal verb choice	0.317	0.752	0.373	0.282	0.457	0.329	0.375	0.618	0.730	0.470
Modal verb placement	0.341	0.761	0.389	0.304	0.486	0.340	0.406	0.647	0.735	0.631
Negation nicht	0.146	0.032	0.203	0.196	0.029	0.113	0.094	0.5	0.613	0.584
Negation kein	0.025	0.258	0.288	0.261	0.114	0.031	0.313	0.206	0.387	0.362
Adverb placement	0.561	0.387	0.627	0.739	0.571	0.103	0.701	0.407	0.516	0.636
Auxiliary past tense	0.219	0.774	0.458	0.478	0.200	0.546	0.563	0.794	0.680	0.768
Participle past tense	0.244	0.806	0.475	0.500	0.229	0.557	0.594	0.824	0.685	0.580

With two exceptions (agreement and independent clauses), the range between the earliest and latest occurrence of a feature in the textbooks can be considered as fairly high with values between 0.362 for negation kein and 0.768 for auxiliary choice in the past tense. The remaining ranges are distributed equally with two to three features in every 100th segment: modal verb choice and separable prefix-verbs in the 400th segment with values of 0.470 and 0.440; order dependent clauses, participle choice past tense and negation nicht follow with 0.573, 0.580 and 0.584 in the 500th segment; in the 600th-segment we find modal verb placement with 0.631, adverb placement with 0.636 as well as dative with 0.629; finally, the 700-values are accusative with 0.741 and auxiliary choice in the past tense with 0.768.

By looking at the ranges for individual features, a better understanding of the data's variability can be reached. These numbers show that the overall variability is fairly high, with almost ever 100th segment being 'taken' by about two to three features.

As could already be expected from the relatively high range, which automatically implies a high degree of variability, the differences between mean and median are, for some features, rather large:

Table 5-3: Mean and median of order, normalized values

Grammatical feature	Normalized Order Median	Normalized Order Mean
Word order independent clauses	**0.049**	0.008
Subject verb agreement	**0.057**	0.069
Negation nicht	**0.146**	0.214
Negation kein	**0.258**	0.209
Accusative	**0.271**	0.286
Modal verb choice	**0.375**	0.470
Modal verb placement	**0.406**	0.490
Separable prefix-verbs	**0.413**	0.422
Word order dependent clauses	**0.538**	0.558
Auxiliary choice	**0.546**	0.559
Participle choice	**0.557**	0.546
Adverb placement	**0.561**	0.512
Dative	**0.639**	0.531

This was another reason that led to deciding against using the mean as the basis for identifying a center, and thus, an order of the introduced features; instead, the median was used. Unlike the mean, the median is not sensitive to outliers, as was mentioned before, which are to be expected in a data set with a range and a high variability.

According to these values, word order in independent clauses and subject verb agreement are – with a relatively large gap between them and the following features – the first two features to be introduced in textbooks, followed by the negation nicht. The features that are introduced the latest are auxiliary choice and participle choice, adverb placement and the dative case.

The differences between the individual median values are not very high and some features can be grouped together, as a quick look at Table 5-3 already reveals: negation kein with 0.258 and accusative with 0.271 are very close in terms of their medians, which means that they are located close to each other in the typical order of features in textbooks. The same is true for modal verb choice with 0.375, modal verb

placement with 0.406, and separable prefix-verbs with 0.413 and, finally, word order dependent clauses (0.538) and auxiliary and participle choice (0.546 and 0.557) as well as adverb placement (0.561). With a median of 0.639, the difference between dative and these four features is slightly larger.

As has been pointed out several times, the range for the different features is quite wide with values up to 0.768. Nevertheless, when looking at the values below and above the median (see Table 5-2), certain patterns or groupings can be identified, which – when combined – can be considered to be a typical order. Table 5-4 is an alternative representation of Table 5-2. It shows how many textbooks introduce the feature in each of the 10 percentiles with 0 indicating the very beginning of the textbook and 1 indicating the very end. Each dot represents one textbook. The framed sections represent in which percentile (s) the median can be found.

Table 5-4: Distribution of point of introduction in nine GaF textbooks, sorted by feature

	0-0.100	0.101-0.200	0.201 - 0.300	0.301 - 0.400	0.401 - 0.500	0.501 - 0.600	0.601 - 0.700	0.701 - 0.800	0.801 - 0.900	0.901 - 1.00
Agreement	ooooooo	oo								
Order independent. clauses	ooooooo		o		o					
Order dependent clauses			o	o	oo		o	o	oo	
Accusative	oo	o	oo	oo		o o				
Dative		oo			o	o	ooo	oo		
Separable prefix-verbs			oo	oo	ooo	o	o			
Modal verb choice			o	oooo	o		o	oo		
Modal verb place.				oooo	oo		o	oo		
Negation nicht	ooo	ooo	o		o		o			
Negation kein	oo	o	oooo	oo						
Adverb placement		o		o	o	ooo	o	oo		
Auxiliary past tense		o	o		oo	oo		oo		
Participle past tense			oo		oo	oo			ooo	

In Table 5-4, a general trend can be observed. When dividing the table into two halves (0-0.500 = early, and .501-1.00 = late), more overlap exists for the earlier sections (79 circles versus 31, respectively) than for the later ones.

In addition to the general trend, specific patterns seem to exist as far as the distribution of the features, namely the point of introduction, is concerned:

1. Features that are taught first: agreement and order independent clauses are predominantly taught at the very beginning of a textbook
2. Features that are taught relatively early: negation, accusative
3. Features that are generally taught at midpoint: modal verb choice, modal verb placement, separable prefix-verbs, adverb placement
4. Features that are taught relatively late: dative
5. Features that are only taught late: none
6. Features that are spread out across positions: order dependent clauses, auxiliary past tense, participle past tense

Sequencing and acquisitional patterns

Before the identified patterns will be compared with findings from research on acquisitional sequencing, Table 5-5 presents these patterns in the form of a hierarchy of difficulty based on the core idea of the acquisitional perspective to difficulty according to Collins et al. (2009): easy structures are acquired early and difficult ones are acquired late. Distances between the categories are only symbolical in nature and do not represent actual differences between the individual values:

Table 5-5: Level of difficulty according to point of introduction

Level of Difficulty	Feature
Easy	Agreement, Order independent clauses Negation, Accusative Modal verb choice, Modal verb placement, Separable prefix-verbs, Adverb placement
Difficult	Dative

Features that cannot be categorized as they are spread out: order dependent clauses, auxiliary past tense, participle past tense.

According to this hierarchy, agreement, and order independent clauses are the easiest ones and are acquired the earliest, followed by negation, whereas dative is the most difficult one and is acquired the latest. The remaining features are generally taught midpoint. In order to assess whether there is a connection between the sequencing in textbooks that was identified in this study and acquisitional patterns in research, and thereby to answer the question whether these patterns may influence the sequencing, findings from the literature review above will be compared to the findings of this study (Table 5-6).

Table 5-6: Comparison point of introduction this study and research findings on acquisitional orders

Structure (and point of introduction this study)	Congruence with literature findings	Literature
Accusative is acquired before Dative	yes	Diehl et al. 2000
Dependent clauses are acquired later than independent clauses	yes	Pienemann 1984
Agreement is introduced as one of the first grammar items	Inconsistent	Maijala 2010, Tammenga-Helmantel 2012 (yes) Aguado 2012 (no)
Modal verbs were introduced prior to the perfect tense	Yes	Maijala 2010, Tammenga-Helmantel 2012
Separable prefix-verbs are mostly introduced at the same time as modal verbs	No	Diehl et al. 2000, Aguado 2012: Separable prefix-verbs are acquired before modal verbs
No patterns could be observed for past tense and separable prefix-verbs	No	Diehl et al. 2000, Aguado 2012: Separable prefix verbs are acquired before the perfect tense
Tense introduction: present, perfect, past	Yes	Tammenga-Helmantel 2012

As can be seen from this table, the dative is acquired later than the accusative (Diehl et al., 2000) and dependent clauses are acquired later than independent ones (Pienemann, 1984), which is in line with the findings of this study. However, no general point of introduction for dependent clauses could be observed in the textbook sequences that were analysed since the introduction was spread out across the textbooks.

More consistency with both research that was to identify acquisitional patterns and research to confirm these patterns in different learning contexts, could be observed in terms of verbal constructions: *haben* (to have) or *sein* (to be) were one of the first verbs to be introduced and conjugated (cf. Tammenga-Helmantel, 2012; Maijala, 2010). While this was not explicitly mentioned in the data analysis, the feature

agreement very often included these two verbs. Furthermore, modal verbs were introduced prior to the perfect tense and, in terms of tenses, the order that was observed here, present-perfect-past is also in line with previous findings (Tammenga-Helmantel, 2012). However, it has to be mentioned that other research has not identified such overlap with acquisitional research for the introduction of tenses of textbooks for German as a foreign language (Aguado, 2012). Thus, the overlap that was identified here should not be seen as a general observation for GFL textbooks.

No consistency could be found for the introduction of separable prefix-verbs. While research suggests (e.g. Diehl et al., 2000) that separable prefix-verbs are acquired before modal verbs and the perfect tense, which has been confirmed by research on textbook sequencing (Aguado, 2012), this does not hold true in this study. Separable prefix-verbs are introduced at roughly the same time as modal verbs, sometimes even later than them.

Overall, the level of congruence between sequencing identified in this study and the patterns identified in research can be summarized as inconsistent.

Synthetic v. analytic sequencing

In order to assess if sequencing decisions are based on synthetic or analytic considerations, the principles formulated by Wilkins (1976) will be used to assess the identified patterns (where possible).

As was already mentioned in the description of the textbooks, they all describe themselves as communicative in nature, which is confirmed when looking at the table of contents of the textbooks. With no exception, the units (and thereby the sequence) are presented based on different topics or communicative tasks. While most textbooks also mention which grammatical structure they will cover in a specific unit, the topics or communicative contexts are in the focus of how the units are presented. Table 5-7

presents an overview of the topics / unit names of the grammatical structures that will be assessed.

As can be seen, none of the chapter topics refers to grammatical structures but exclusively to communicative contexts. For each individual textbook, Wilkin's operationalization of analytic syllabi seems to apply, which are "organized in terms of the purposes for which people are learning and the kinds of language performance that are necessary to meet those purposes" (Wilkins, 1976, p.13).

However, while the diversity of topics is not very high, which results in an occasional overlap, patterns for a connection between the introduction of the structures of interest in this study and specific communicative topics could not be observed. Exceptions applied to agreement and word order in independent clauses. Since these are the earliest structures to be introduced, the communicative topics are usually very similar and cover scenarios such as introducing oneself, talking about hobbies, the weather other short conversations in settings such as cafés etc.

Table 5-7: Topics used for introduction of feature, sorted by textbook

	Wie geht's?	Kontakte	Treffpunkt Deutsch	Berliner Platz	Sag Mal	Deutsch – Na klar!	Schritte	Studio D	Menschen
Accusative case	Shopping and store hours	Possessions and leisure time	Family life	Going grocery shop.	At school	Holidays	Leisure time activities	Talking about your apartment	In the office: phone conversations
Agreement	Family, countries, lang.	Names	Expressing states and actions	Introducing oneself	Hello! How are you doing?	Hello and welcome!	Introducing oneself	Conversations at a café	Introducing oneself
Dative case	Eating in and out	Money and work	Holidays and vacation	Getting around in Berlin	Holidays	Eating and drinking	Gifts and celebrations	Getting around in a city	City life: talking about a city
Dep. clauses	Holidays and vacation	Talents, plans, duties	Talking about food	n.a.	City life	Daily routines	Talking about travel plans	Talking about learning a language	Health and fitness
Indep. clauses	Family, countries, lang.	Who I am and what I do	Talking about the weather	Introducing oneself	At school	Seasons and the weather	Introducing oneself	Conversations at a café	Making plans with friends
Modal verbs	In the city: spotlight on Austria	Talents, plans, duties	Everday-life / routines	Finding an apartment	Family and friends	Going to the movie theatre	Lifelong learning: talking about skills	Profession and every-day life	Leisure time activities
Past tense	Holidays and vacation	Events and memories	Looking back: describing past events	Talking about the past	Holidays	Traveling and transportation	Lifelong learning: talking about skills	Holidays and vacation	Talking about routine and events in the past
Sep. prefix-verbs	Exchange offices and credit cards	Who I am and what I do	Everday life / routines	Talking about routine	Food	Going out	Talking about a daily routine	Scheduling events	Travel and transportation

In addition to examining the chapter topics, the identified patterns are assessed based on two of the three criteria for synthetic syllabus design formulated by Wilkins (1976): criterion 2: *simple structures before complex structures* and criterion 3: *interaction between linguistic structures*. Criterion 1 (*new structures refer back to old ones*) was impossible to assess since only patterns and not entire sequences are compared.

Criterion 2 and 3 can only be assessed when related structures are in the focus, such as accusative and dative or dependent and independent clauses. When the introduction of modal verbs is contrasted with the perfect tense or separable prefix-verbs, these criteria cannot be assessed because contrasting these specific structures was based on the fact that a set of fixed structures was chosen for analysis.

Decisions about the complexity of a structure were mainly based on how many steps are necessary to arrive at the correct form (Housen & Simoens, 2016).

Table 5-8: Sequential patterns and criteria for syntactic syllabi (Wilkins, 1976)

Sequential pattern	Simple structures before complex structures	Interaction between linguistic structures
Accusative is acquired before Dative	x	x
Dependent clauses are acquired later than independent clauses	x	x
Agreement is introduced as one of the first grammar items	x	n.a.
Modal verbs were introduced prior to the perfect tense	x	n.a
Separable prefix-verbs are mostly introduced at the same time as modal verbs	x	n.a.
Tense introduction: present, perfect, past	x	x

As can be seen in Table 5-8, the structures that allowed for a comparison with the features of a synthetic syllabi all meet criteria 2 and 3.

Discussion

As was demonstrated, several sequential patterns have emerged, which can be considered to be a typical sequence across GaF textbooks, with more overlap in the early introduction phases of the textbook than was true for the later sections. Following the definition of difficulty within the acquisitional perspective (Collins et al., 2009), this may suggest that there is more agreement about what is easy than about what is difficult. With a larger sample size, these patterns may become even more evident.

With regard to whether sequential decisions seem to be determined by analytic or synthetic considerations, one has to conclude that both holds true to a certain extent. The

layout of the textbooks, the organization of contents around communicative contexts and the self-description as mainly pursuing communicative objects lead to believe that sequential decisions are predominantly analytic in nature. All textbooks seemed to be "organized in terms of the purposes for which people are learning and the kind of language performance that are necessary to meet those purposes" (Wilkens, 1976, p. 13). While this holds true for individual textbooks, such organization, as would have been manifest in the form of using similar topics for similar structures, could not be identified for the patterns that were in the focus here.

The (very limited) analysis showed that most criteria related to synthetic syllabi were fully met. However, not only was the analysis very limited (since no overall syllabus could be analysed in depth), the two criteria *simplicity before complexity* and *interaction* are not exclusively a manifestation of synthetic syllabus design but can be considered to be an aspect of textbook design that follows an explicit decision-making process. Presumably, these decisions were based on tacit assumptions about L2 learning difficulty in general and not necessarily with the aim to structurally control the gradation process.

Furthermore, the patterns that could be observed here are – to a certain degree – congruent with research on acquisitional patterns. However, as was pointed out, consistency was not observed across studies from other research and also not for all structures and textbooks that were in the focus of this project. Since no information is provided by the authors on how sequencing decisions are made, one can only speculate about the reasons and motivations behind textbook sequencing.

With the vast number of studies on acquisitional patterns or aspects such as learnability or teachability (e.g. Pienemann, 1985; Krashen & Seliger, 1975), one may be

quick to conclude that it seems to be logical and rather easy to make sequencing decisions in textbooks based on these research findings. However, not only do textbook authors have to deal with various challenges and constraints, the connection between research and theory is not as straightforward as one may assume.

First of all, textbook authors need to tailor to a very heterogenous and dynamic target group that consists of the learners (including their parents, depending on the context), the teachers and the institutions. The needs and demands of these groups spread out to various different areas for consideration, such as practical concerns (pricing and availability/distribution of the book), learning-related issues (e.g. SLA research but also learner and teacher perceptions, identification of *who* the (typical) learner is, learner types), teaching-related issues (congruence between institutional curricula and textbook contents, learning and teaching objectives, teaching style, etc.), and aspects that are relevant to the publishing market that is defined by consistent “power [struggles] and hierarchies” (e.g. competition, online retailing or RPM (Retail Price Maintenance (Trentacost & Pilcher, 2021)). With regard to grammar books alone, Byrd (1995) has identified 10 “Design Features” that need to be considered in the process of textbook design. Most of these are tailored to the diverse needs of the target groups mentioned above in addition to the requirements formulated by the language in question.

Even if one were to be able to consume all relevant research in the fields described above (which is an almost impossible task by itself), the cycle of textbook publishing is not in-sync with research cycles, which are usually shorter, which means research is published faster and adds findings more frequently to the existing body of research than is true for textbooks, which usually remain the same for several years (e.g. because of

the costs associated with producing a revised edition of the textbooks). Furthermore, research findings may contradict themselves, which gives rise to further research and another research cycle.

In addition, as was stated above, the connection between research and theory or practice is not a simple one-way form of informational transfer, namely from research to practice. Stern (1983) has described these interactions in his *General Model for Second Language Learning*. According to this model, practice is one out of three levels: before the practice Level (3) has been reached, Level 1 forms the foundations, for example, the fields of Linguistics, Sociology Psychology or Educational Theory. These components are all in a mutual relationship with both each other and with the components of Level 2 (Interlevel), which is comprised of Learning, Language and Teaching and can be summarized as “Educational Linguistics Theory and Research”. This level is directly connected with Level 3: Practice. Textbooks or “Materials” are only one out of 5 aspects of what he calls “Methodology”; this section is not only in interaction with Level 2 but also with the other component in Level 3 “Organization”. The latter consists of institutional aspects such as “Planning and administration”, “Higher education” or “Teacher education” (p. 338). Not only does this model show that multiple components are part of the supposedly simple relationship between research and practice, it also shows the numerous ways of how and where information is exchanged. Therefore, one cannot expect textbook authors to directly consume, transfer and implement research findings for textbook design.

It is therefore not surprising that, according to Byrd, most decisions related to textbook design are “based on traditions about grammar materials and their organization

rather than on careful rethinking of either the content or its organization" (p. 46). Relying on tradition is not only convenient, it also accounts for the "conservatism of both students and teachers about appropriate content and activities in language learning courses" (ebd. p. 46). This finding is in line with what Reid (1995) suggests in her Materials writer's guide: "The golden rule for any commercially published materials is that the basic idea must be 20% unique and 80% familiar in order to sell well" (p. 66).

Conclusion

This study aimed to identify an overall typical order of grammatical sequences in textbooks for German as a foreign language. While various studies exist that compared textbook sequencing of specific structures with acquisitional patterns, they were all lacking a feasible way of expressing relativity of the positions they were discussing. In this study, normalization was used successfully to express such relativity both within and across textbooks to identify a typical order. This method will hopefully facilitate future research on textbook sequencing.

In addition to this methodological concern, looking at sequencing in textbooks for foreign language learning is also a topic of practical concern that lends itself to comparisons with research: not only is the research body on acquisitional sequences very comprehensive, the data that is gathered from both contexts, that is, sequences in textbooks and acquisitional sequences, is directly comparable due to the similarity. As a result of this proximity, the discussion revolving around the connection between research and practice is unavoidable, which is why it was also the focus of the preceding discussion section of this study. In many studies, a gap between research and practice is identified

which often leads to a call for bridging this gap in the concluding sections of research articles; often with the subtext that practitioners should be the ones more interested and more involved in research findings and how they can be implemented in practical settings. Maybe it is now time for researchers to become more aware of practical constraints, for example, in textbook design, instead of holding teachers alone accountable for bridging this gap.

Creating more awareness of the complexity of the relationship between practice and research would be a first step towards more fruitful cooperation. When looking back at Stern's (1983) model, this awareness has to include the understanding that all levels and components involved in SL teaching are equal, which means one should not be valued higher than the other. Following this model, a dialogue between practitioners and researchers has to be mutual and it should also happen in and between all three levels.

Finally, the parties involved should not only be aware of the constraints involved in textbook design but, to some degree, of the fact that these constraints have to be accepted. For these reasons, this study does not conclude with a simple call for bridging the gap between research and practice. Instead, it is supposed to encourage a more critical and realistic assessment of the relationship between research and practice, and the (sometimes unrealistic) wish for (rapid and immediate) innovation.

Chapter 6: Conclusion to the Dissertation

This final chapter offers a general discussion of the set of empirical studies presented here; it includes a summary of the research and comparison/synthesis of the findings from the four studies. This is followed by consideration of the contributions and pedagogical implications of this enquiry as well as its limitations and possible next steps for investigation of this topic.

Introduction

In the four studies presented in this dissertation, I have examined from different theoretical perspectives the notion of learning difficulty with respect to German grammar in the context of foreign language study. The first perspective was student orientation, that is, learners were asked to rate the perceived difficulty of certain features and provide reasons for their choice. In the second study, difficulty was approached from a psycholinguistic perspective: it was examined whether difficulty needs to be defined based on the distinction between implicit or explicit language knowledge. The third study examined the metalanguage that is used in beginner GFL books and how it may be connected with the concept of difficulty. In the fourth study, the question addressed was whether there is a typical sequence across beginner GFL textbooks in which grammatical structures are introduced and which factors may contribute to this order. The connection between Study 4 and the concept of difficulty was established by contextualizing the findings within the acquisitional approach to difficulty (Collins et al. 2009), which defines easy structures as those that are acquired early and difficult ones as those that are acquired late.

It is assumed that a comprehensive understanding of the concept can make a useful contribution to pedagogical decision-making, especially for learners of German at the beginner level studying in post-secondary level contexts in English-speaking countries. This final chapter offers a general discussion of the four studies presented here, which is followed by consideration of the contributions and pedagogical implications of this enquiry as well as its limitations and possible next steps for investigation of this topic.

Overview of the four studies

Each study addressed a different perspective on the topic of difficulty and necessarily involved different types of data and different analyses. Table 6-1 provides an overview of the focus, the data collected and the type of analysis carried out for each of the studies. A brief summary of each study then follows.

Table 6-1: Overview of the four studies

Study	Perspective/focus	Data	Analysis
1	*Learner perceptions of grammatical difficulty*	Writing prompt Ratings of difficulty on Likert scale Open-ended questions about reasons for perceived difficulty	Accuracy rates from writing samples Mean ratings Correlation between accuracy rates and perceptions Themes in learners' responses to open-ended questions
2	*Grammatical difficulty and the dimensions of implicit and explicit knowledge*	Implicit knowledge: Oral Imitations Timed grammaticality judgements Explicit knowledge: Untimed grammaticality judgements Metalinguistic knowledge	Mean test scores T-tests between the different measures and different learner groups
3	*Metalanguage in beginner textbooks of German as a foreign language as a source of difficulty*	Metalanguage in grammar explanations in 6 GFL textbooks	Frequency of different features: explained vs assumed terms; transparent vs opaque terms; number of distinct metalinguistic terms; explanation strength
4	*The sequencing of grammatical features in beginner textbooks of German as a foreign language and grammatical difficulty*	Grammar topics in 9 GFL textbooks	Creation of an "average" sequence of grammar topics across textbooks

Study 1: Exploring Learners' Perception of Grammatical Difficulty in the German as a Foreign Language Classroom

In the first study of this dissertation, the questions what students find difficult about learning German and, more importantly, for what reasons were answered. Previous research (Chavez 2016, 2017) had identified eight categories in learners` explanations for why certain grammatical structures were difficult. With 27% and 25% respectively, memorization and complexity of rules account for the largest number of reasons for perceived grammatical difficulty, followed by speed of processing with 13%. This paper reports findings from a partial replication of Chavez' studies. The data collection and the identification of the features in question were partly based on Chavez´ findings from 2016 and partly on accuracy rates computed from writing samples that were collected from 112 beginner and intermediate learners (A1 to B1) of German at a university in Canada. Eliciting data in the form of writing samples with a task that was appropriate for the learners in terms of content, difficulty level and type of task allowed for an analysis of data that was relevant to the learner group in question. It further served to confirm the findings of Chavez` study from 2016.

Based on these writing samples, a survey was designed that required the participants to rate the difficulty of the structures on a scale from 1 to 6. In a second step, they had to choose three structures that they considered to be representative of being *not difficult at all*, *moderately difficult* and *extremely difficult* and provide reasons for their choice. Reasons were provided in an open format to make sure that the responses were not biased by having to choose from pre-formulated answers. Themes were identified in the thus elicited data.

A correlation coefficient was computed to find out whether there is a connection between the mean values for the perceived level of difficulty and accuracy rates. With very few exceptions, there does not seem to be an association between these two.

The themes identified in the qualitative data confirm those of Chavez' study (2017), but add other dimensions to our understanding of the learner's view of what is difficult about studying German, e.g. instruction, cognition and rule consistency. For structures to be perceived as easy, memorization, instruction, and cross-linguistic transfer were identified to be contributors to the easiness of these structures. For difficult structures, the most frequently identified ones were memorization, rule complexity, cross-linguistic transfer and cognition.

In contrast to Chavez' reluctance to draw pedagogical implications, it is argued here that this study provides pedagogically useful information about learners' understanding of the individual structures and how they should be taught; for example, increasing the amount of explicit grammar instruction and carefully assessing the use of the target language in the classroom.

Study 2: Grammatical Difficulty in German and the Dimensions of Implicit and Explicit Knowledge

This study answered the question what grammatical structures are easy/difficult for English-speaking learners of German on measures of explicit and explicit knowledge. It also aimed to assess whether these structures are the same for different proficiency levels, namely beginner and intermediate level learners (A1 to B1).

Four instruments were used to measure performance: the oral imitation test and the timed grammaticality judgement test for implicit knowledge and the metalinguistic

knowledge test as well as the untimed grammaticality judgement test for explicit knowledge. The decision to choose these tests was based on previous research by Ellis 2006 in a similar study that looked at learners of English. The scores of these tests were translated to different levels of difficulty: *easy*, *moderately difficult* and *difficult*.

The results showed that, overall, more structures were easy (six out of eleven) on measures of implicit knowledge than was the case for explicit knowledge (five out of eleven), which was surprising since implicit knowledge is harder to attain than explicit knowledge. On the other hand, these findings may be in line with the instructional approach of the learners' German classes, which puts very little emphasis on explicit grammar teaching.

It was further found that there was very little consistency for which structures were either *easy* or *difficult* on both measures, which confirms that difficulty needs to be approached based on the distinction between explicit and implicit knowledge since the mental processes involved in performing these tasks as well as the factors that contribute to the difficulty for each knowledge type are very different.

Findings for the different proficiency groups were, with two exceptions, not surprising: the level of difficulty decreased for most structures on both measures; however, differences between the mean values were not statistically significant.

Based on these results, it can be concluded that the distinction between the different knowledge types should not only be kept in mind in future research on the concept of difficulty, but it also gives reason to believe that learners would benefit from instructional material that is designed with this distinction in mind.

Study 3: Metalanguage in German as a Foreign Language Textbooks as a Source of Learning Difficulty

Study 3 focused on the use of metalanguage in the explanation of 13 grammatical structures in six beginner textbooks for German as a foreign language and how they may be connected with the concept of difficulty. Metalanguage was operationalized in four different ways: 1) the overall number of metalinguistic terms used for a feature, 2) the number of distinct metalinguistic terms, 3) the number of explained versus assumed terms as well as 4) the number of transparent and opaque terms (Berry 2010).

Numerous common features of the metalanguage across all textbooks were found: the number of opaque terms is higher than the number of transparent terms and there are consistently more explained than assumed terms although the latter make up about 30 %. Furthermore, terms were often not explained when they first occurred in the textbook but in later chapters. While no overall connection between assumed/explained terminology and opacity/transparency could be observed, two patterns were identified for morphology and syntax: for morphology-related terms, high values for both opacity and explained terminology were observed while for the domain of syntax, high values for transparent and explained terminology occurred.

The findings violate the rules for reducing the difficulty of metalanguage as suggested by Heringer and Keller-Bauer (1984) due to 1) a lack of adaptivity to the target group and 2) potentially missing comprehensibility of the explanation of a term.

Another common finding was that opacity was mostly due to the fact that the terminology was taken from Latin-based grammars with no explanations or translations provided. Thus, a great opportunity is missed to increase the level of transparency that is

used in the grammar sections and, at the same time, the chance to reduce the discrepancy between traditional approaches to language and language learning, as manifested in the predominant use of Latin-based terminology (Funk 1995), and the communicative approach that most textbooks claim to pursue.

This Latin-based approach can be seen as rather Eurocentric, since the implied perceptions of learners as possessing explicit language or linguistic knowledge are more likely to be the result of European syllabi for secondary schools than would be the case in North-American high school curricula.

Study 4: The Sequencing of Grammatical Features in GFL Textbooks

This study aimed to answer the question whether there is a typical sequence of grammatical structures in textbooks for German as a foreign language at the beginner level. It further served to identify aspects which may have an influence on this sequence, if said typical sequence can be identified. These questions were answered by analyzing nine beginner textbooks that were either monolingual in German or bilingual in English.

With regards to the concept of difficulty, sequencing is understood as part of the acquisitional approach to the concept, which defines structures that are acquired early as easy and those that are acquired late as difficult (Collins et al. 2009). Sequencing is further strongly connected with research on L2 acquisitional sequences. Therefore, it was analysed whether acquisitional patterns for German as a foreign language may be one of the factors to influence textbook sequencing. Due to the strong communicative focus of the textbooks, which emphasizes the importance of the learners' capability to act in different communicative settings, the topics within which the grammatical structures were

introduced were assessed to find out whether sequencing decisions may be analytic or synthetic in nature (Wilkins, 1976).

Looking at sequences across different textbooks poses the challenge to find a way to make the data, that is, the different positions and their relativity within and across the textbooks, comparable. Since previous research had only looked at individual structures or patterns and not overall sequences, such method did not exist. This study introduced the concept of normalization to fill this gap in research and to offer a methodological approach for future research on the topic of sequencing.

Findings indicate that several patterns exist across German as a foreign language textbooks, which – when combined – can be considered to be a typical sequence. More overlap across textbooks existed in the early introduction phase than was true for later sections, which may suggest there is more agreement about what is easy than about what is difficult.

While many of these patterns are congruent with research on acquisitional sequences (e.g. positions of accusative versus dative, dependent clauses and independent clauses, tense introduction), no patterns could be identified with regards to the topics within which the structures were introduced.

The high similarity between research findings and how they are implemented in textbooks gave rise to a discussion about the relationship between practice and research, which is often addressed in the form of a simple call to bridge the gap. This study concludes with a call for a more realistic assessment of the complexity involved in this relationship and what the implications are for textbook design.

Comparisons of the order of difficulty in Studies 1, 2 and 4

How similar are the orders of difficulty that have emerged from the different perspectives explored in Studies 1, 2 and 4? Table 6-2 presents the ranks from easiest (1) to most difficult (11). The ranks for learner perceptions are based on mean values that were computed for data from a Likert scale (1-6) in a questionnaire; ranks for implicit and explicit knowledge are mean values for learner performance on four measures of these two knowledge types; the hierarchy for textbook sequencing is based on the normalized order of median values. When several features have the same rank, the values in the studies were identical. Ranks are only provided for features that were assessed in all four studies.

Table 6-2: Rank orders of difficulty for learner perceptions, knowledge types and textbook sequencing

Grammar feature	Learner Perceptions		Knowledge Types		Textbook Sequencing
	100	200	Implicit	Explicit	
Accusative	5	7	8	7	4
Agreement	1	2	3	4	2
Auxiliary choice past tense	6	2	9	8	8
Dative	8	9	10	5	11
Modal verb placement	4	4	5	2	5
Negation	4	6	2	9	3
Participle choice past tense	9	8	6	10	9
Separable prefix verbs	2	5	7	3	6
Adverb placement	3	1	5	3	10
Word order: Dependent clauses	7	3	1	1	7
Word order: Independent clauses	7	3	4	6	1

In order to compare the ranks from the different theoretical perspectives, the ranks based on the individual studies can be further grouped into categories. Table 6-3 indicates whether these structures are *easy* (ranks 1-4), *moderately difficult* (ranks 5-8) or *difficult* (9-11). Consistency can be assessed based on 1) the overall overlap in the difficulty classification between approaches, and 2) the distance between the different levels of difficulty; for example, are they one category apart: easy and moderate, or two categories: easy and difficult.

Before the comparisons are presented, the following considerations need to be kept in mind:

1. The three levels of difficulty are an approximation to make the findings comparable and they do not have the same reliability as a statistic measure of association would have.
2. Consequently, an overlap does not say anything about its strength, consistency or its direction.
3. Due to the difference in scaling, the categories assigned to the different structures are not always identical with the ones that were assigned in the studies (this primarily refers to knowledge types, where such categorization was part of the study).
4. Data from the different approaches was not gathered from the same participants across studies and so specific observations cannot be paired. While this was one of the reasons for a lack of statistical analysis, it also needs to be taken into account when interpreting the findings.

Table 6-3: Levels of difficulty: learner perceptions, knowledge types, textbook sequencing

Grammar feature	Learner Perceptions		Knowledge Types		Textbook Sequencing
	100	200	Implicit	Explicit	
Accusative	Mod	Mod	Mod	Mod	Easy
Agreement	Easy	Easy	Easy	Easy	Easy
Auxiliary choice past tense	Mod	Easy	Diff	Mod	Mod
Dative	Mod	Diff	Diff	Mod	Diff
Modal verb placement	Easy	Easy	Mod	Easy	Mod
Negation	Easy	Mod	Easy	Diff	Easy
Participle choice past tense	Diff	Mod	Mod	Diff	Diff
Separable prefix-verbs	Easy	Mod	Mod	Easy	Mod
Adverb placement	Easy	Easy	Mod	Easy	Diff
Word order: Dependent clauses	Mod	Easy	Easy	Easy	Mod
Word order: Independent clauses	Mod	Easy	Easy	Mod	Easy

Table 6-3 reveals that agreement is the only feature where there is complete consistency across the studies. Agreement appears to be easy. At the opposite end of the scale, dative and participle choice each exhibit three classifications as difficult and two as moderately difficult. Given this overlap, these two features appear to be difficult for L2 learners. The accusative has an overlap of four moderately difficult classifications but an easy classification based on the textbook analysis.

For the remaining structures, no such clear overlap across the different data-sets can be identified; presumably because 1) the different approaches are not represented equally, that is, for textbook sequence, only one category exists while there are two each for learner perceptions and knowledge types. 2) the three corresponding studies have

shown that differences exist between learner groups (at least to a certain degree for learner perceptions) and knowledge types, which would make a complete overlap implausible.

A partial overlap can be identified for the following structures: for modal verb placement, learner perceptions and explicit knowledge overlapped as well as implicit knowledge and textbook sequencing; for word order independent clauses, an overlap between implicit knowledge and textbook sequencing is evident as well, while explicit knowledge matched with the perceptions of the 100-level learners.

For word order dependent clauses, no overlap seems to exist for knowledge types and textbook sequencing while it does seem to exist for knowledge types and the perceptions of the 200-level learners. Partial overlap across three approaches was also observed for separable prefix-verbs: the perceptions of the 200-level learners, implicit knowledge as well as textbook sequencing.

While partial overlap is evident for the three remaining structures adverb placement, negation and auxiliary choice, consistency in terms of the distance between levels of difficulty across the data-sets was lower than for the other eight structures: for adverb placement, the level of difficulty for learner perceptions is easy while in textbook sequencing it is difficult. For auxiliary choice in past tense, the levels of difficulty also range between easy, moderately difficult and difficult.

In addition to these specific results, the following observations can be made on a more general level:

- For 8 out of 11 structures, full or partial consistency across the different approaches could be observed, which can be considered to be very high.
- Overall, when comparing perceptions and textbook sequencing, learner perceptions are either identical with or lower than the level of difficulty identified for sequencing.
- When comparing knowledge types and textbook sequencing, 9 out 11 structures were less difficult or equally difficult for explicit knowledge and sequencing; for implicit knowledge 6 structures were equally difficult, 2 were more difficult and three were less difficult.
- For the comparison between knowledge types and learner perceptions, the following was observed: between 100-level learners' perceptions and explicit knowledge, there was an overlap for 9 out of 11 structures; between 200-level learners' perceptions and explicit knowledge, the overlap was 5 out 11 structures. With regard to implicit knowledge, the overlap for the 100-level learners was 3 out of 11, for the 200-level learners it was 7. For the structures where no match was identified, no patterns could be found in terms of whether the differences in the level of difficulty were – overall - towards a higher or lower difficulty level.

For the purpose of this dissertation, the finding that there is – overall – a relatively high consistency across the different approaches to difficulty gives reason to be optimistic since it makes the design and implementation of measures to overcome difficulty less challenging. However, given the differences in nature of these approaches with one relying on subjective perceptions, the other on performance on measures for different

knowledge types, and the analysis of instructional materials, this overlap is also rather surprising.

Knowing that structures are equally difficult or less difficult in the subjective approach to difficulty (i.e. learner perceptions) and textbook sequencing is reassuring on the one hand because one may argue that sequencing decisions in textbooks do not negatively influence learner perceptions. On the other hand, it raises more questions in terms of the type and the direction of the relationship: are learners' perceptions of difficulty a logical consequence of the sequencing decisions? For example, is something that is introduced later on in the textbook perceived as easier because the level of exposure to the language has been higher at this point than was the case for structures that were introduced earlier?

While no association was observed between performance (operationalized as accuracy in writing samples) and the perception of difficulty in Study 1, the high degree of overlap between the perception of 100-level learners and performance on measures of explicit knowledge (9 out of 11) is noteworthy. With the fact in mind that these findings stem from different learner groups and that this was not a statistical analysis, one can still speculate that 100-level learners may be more in-sync with the explicit dimensions of language learning since there was not much time for implicit knowledge to develop in beginner classes. Furthermore, many forms of assessment target explicit language knowledge, even in communicative GFL settings. The fact that the overlap between perceptions and implicit knowledge was much lower with a value of 3 for the 100-level learners strengthens this claim.

Finally, the connection between knowledge types in Study 2 and the textbook sequences from Study 4 needs to be evaluated. For explicit knowledge, the level of difficulty for 9 out of 11 structures was either the same or less difficult than for textbook sequencing. As was true for learner perceptions and textbook sequencing, there seems to exist a very high level of congruence, which may be read as a manifestation of the fact that what a textbook typically introduces in a specific position is in-sync with what learners are capable of or ready for processing at that given time (Pienemann, 1985). More fine-grained analyses are needed to support this claim and shed further light on the connection between these two manifestations of learner difficulty.

So far, findings from Study 3, the use of metalanguage in beginner textbooks for German as a foreign language, have not been taken into account because the data did not lend itself to identifying a rank order of difficulty. However, the following observations can be made: the impact of the predominant use of Latin-based terminology as well as the inconsistency in regards to the use of metalanguage were *assumed* to contribute to learning difficulty (mostly against the background of the rules identified by Heringer and Keller-Bauer, 1984). When assessing this assumption in light of a key finding from Study 2, namely the fact that more structures were more difficult on measures of explicit knowledge than on measures of implicit knowledge, there seems to be reason to believe that this assumption may be justified since metalanguage is one of the two major components that contribute to difficulty associated with explicit knowledge (Ellis 2006). The observations made in the comparison of the rank orders of difficulty further showed that 100-level learners seem to be particularly in-sync with the explicit dimension of

language learning, which means that they may be particularly susceptible to metalanguage as a contributor to difficulty.

When looking back at the findings from Study 1, some of the most frequently mentioned themes related to learners' perceptions of the difficulty of a specific structure can be connected to metalanguage, namely: instruction, cross-linguistic transfer and rule complexity. With regard to the latter, using mostly opaque, Latin-based terminology can be seen as a contributing factor to that complexity. When the terms that are used to explain a rule are simplified or made transparent, the perceived complexity of the rule may be positively affected.

Cross-linguistic transfer was also among the most frequently mentioned contributors to difficulty, especially for 100-level learners. It is quite natural for adult learners to compare new linguistic input either with their L1 or with other languages they have previously learned. Thus, a cross-lingual teaching strategy as recommended by Stern (1992) is likely to enhance both grammar learning and increase overall learner-centredness. It may also contribute to a decreased level of difficulty in students' perceptions of a specific structure but, of course, that is an empirical question.

The final theme from Study 1, instruction, was often mentioned in the context of how much or how little a specific structure is practiced; however, instruction is also directly connected with metalanguage as it can be seen as its tool or vehicle. In sum, targeting metalanguage in instruction overall and in textbooks in particular has potential for decreasing difficulty associated with the German language.

Furthermore, there is a similarity between the use of metalanguage and grammatical sequencing in GaF textbooks, which was done in Study 4: not only do both

approaches direct attention to instructional material, they also imply certain assumptions about the learner. Very little is known about the details of how textbook writers go about choosing content and deciding on a sequence for their textbooks. However, it is probably safe to assume that – at the very least – their choices are made against the background to provide material that facilitates the language learning process. While the sequence that was identified in Study 4 is, to a certain extent, in line with research on L2 acquisitional patterns, it is further very similar to learner perceptions and can thus be seen as facilitating to the process of decreasing learning difficulty. This, however, was not true for the use of metalanguage. Not only have the findings shown that the choice of the terms that were used violates rules for avoiding or decreasing difficulty in the use of metalanguage (cf. Heringer & Keller-Bauer, 1984), it was also demonstrated that metalanguage seems to be chosen with a very specific learner group in mind, namely one that has prior knowledge in either other foreign languages or in linguistics. Or, in other words, the learners' skills are likely to be over-estimated (Berry, 1997).

In the following section, the above-mentioned implications about the learners associated with metalanguage will be addressed.

Implications of the findings

The fact that the metalanguage used in beginner textbooks for German as a foreign language implies an over-estimation of the target group is in strong conflict with the depiction of learners in research. Chavez' study from 2017, which was partially replicated in Study 1 of this dissertation, is an example for the other end of the spectrum, namely the under-estimation of the learner: while the conclusion she draws based on her

findings are valid, for the most part, the depiction of the participants of her study, who are all language learners, is very negative and presents them as 'flawed' while dismissing most of the potential that their voices may have. In that regard, Chavez study is representative of a large body of SLA research, in which language learners are often presented in the light of their deficiencies instead of their potential (Ortega, 2014; Holliday, 2006). An over-estimation of the learners in textbooks versus an under-estimation in research creates an area of conflict and may not only bear the risk of less learner-oriented research in the future but it may also lead to designing and implementing instructional material that is tailored to a target group that does not actually exist. Thus, another dimension is created through which the gap between research and practice is increased further instead of decreased. For approaching concepts such as difficulty, which can only be depicted comprehensively by focusing on both practice and research, the consequences of this gap can have a very negative effect on accommodating learner needs.

Limitations of this research

In addition to the limitations that were mentioned for each study, the comparison of difficulty across different data-sets and theoretical perspectives is subject to additional limitations. Although 'translating' the findings from each study into a rank order of difficulty which was then categorized into different levels of difficulty allows for a straightforward comparison of the findings, a statistic analysis would have offered a more precise comparison. Strength and direction of an association are only detectable statistically. However, due to the way my research project unfolded and time constraints, it was not

possible to have the same students complete all of the instruments from Studies 1 and 2. Given the dense curricula in these courses, it was not feasible for the teachers to give up more hours than they already did. Overall, the sample size was rather small; including more than one university would have allowed for a higher number of participants, but it would have also required a very high level of coordination and of teachers' willingness to sacrifice precious class time.

Contributions of this research

By approaching the concept of difficulty from four different perspectives, this dissertation research answers the call for more comprehensive approaches rather than a single perspective, which is – to this day – the predominant research approach to difficulty. In particular, the findings contribute to applied SLA scholarship on German as an L2 that links theory and research to practice and it provides concrete insights that may be useful for teachers and textbook writers.

Methodologically, it has been demonstrated that – while the partial replication of an existing study (i.e., Chavez, 2017) confirmed most findings of the original study – the replication in a different research setting elicited additional data and, in this case, identified more categories for describing reasons for learning difficulties from the learners' perspective.

Finally, Study 4 introduced a new approach to identifying an average teaching sequence across different GFL textbooks, which may be useful for future research on this topic.

Directions for further study

In addition to the suggestions for future research that were provided for each study, the difference between the tacit depiction of the learner in textbooks (i.e., in practice vs. in research, as described above) should be in the focus of future research. Future studies could systematically compare and contrast how learners are depicted in these two contexts including questions on what aspects contribute to these representations, what the consequences are and how to strike a balance between the depictions in these two settings. With regard to difficulty, such research could help to identify more learner-oriented approaches to facilitating learning difficulty in different contexts.

Furthermore, more voices should be included in future research. In addition to the learner's perspective, the perceptions of teachers and textbook designers could be included in assessing different dimensions of difficulty in instructed language learning settings.

Similar approaches to difficulty for different languages and in different settings, such as secondary or even primary level education as well as different instruction and learning styles may serve to 1) validate and 2) extend research findings on the concept of difficulty. Likewise, the consistent inclusion of more learner variables such as pre-existing language knowledge as well as more proficiency levels in addition to the ones that were part of this research, could elicit even more nuanced data on the concept of difficulty.

Finally, the effectiveness and practicability of the pedagogical suggestions that were formulated based on the findings of the four studies of this thesis can be in the focus of future research.

Final comments

The concept of difficulty is, by definition, a highly subjective and context-sensitive topic, which has been demonstrated in this study. Therefore, an assessment of the overall difficulty of the German language as a whole (even if focusing on only one group of L1 learners, namely English-speaking learners of German) would be as valid as anecdotal evidence or the public laymen discourse on the difficulty of the language outlined in the introductory section of this dissertation. Instead, these final comments are devoted to the contributing factors to difficulty that should serve as a starting point to reduce the effect of these factors.

It has been shown that most contributors to difficulty can be targeted directly in the form of instructional decisions: the lack of both practice and learning strategies to help transfer explicit into implicit knowledge as well as too little explicit instruction (as perceived by the language learners), the use of inconsistent metalanguage that is not tailored to the target groups in question, the failure to distinguish between implicit/explicit knowledge when assessing performance and when making decisions to increase the quality of instruction, and, finally, the lack of consistent, research- and practice-driven sequencing decisions in textbooks for German as a foreign language.

These contributors to difficulty could be encountered either by adapting instructional material, usually a long and slow process, or by implementing changes in daily instructional practice, which assigns major responsibility for decreasing learning difficulty to textbook designers and teachers. The fact that the overall levels of difficulty were fairly consistent across the approaches should make systematic intervention more feasible.

While the focus on specific approaches to difficulty and on specific grammatical structures does not allow for an overall assessment of the difficulty of the German language, findings from the studies conducted here may instead serve to encourage research on other languages to find out if the contributors to difficulty apply to these contexts as well.

As was outlined in the introduction to this thesis, enrolment in language courses in general is on the decline and not only with regard to German. Therefore, investigating whether the findings of the four studies that were conducted here apply to other languages as well is much needed, especially after having identified the close connection between instruction and the concept of difficulty and the role it plays for students when deciding whether or not to continue or discontinue a university (language) course (Weseley, 2010).

Due to the pandemic, online teaching has become the new reality in the majority of institutional learning settings worldwide. While all courses and languages are affected by current circumstances, it should not be dismissed as another important factor that may have a negative impact on enrolment, especially when (language) courses are taken on a voluntary basis. It can be expected that enrolment numbers in North America will continue to decrease. With the current travel restrictions in place due to the pandemic, visiting Germany (or any other country) as a tourist or to visit family is more cumbersome than it usually is, which may have a negative effect on students' decisions to learn languages in the near future.

Similar reasons may apply to the field of German as a second language, which so far had not seen a decline in enrolment. Not only has the number of refugees gone down due to the pandemic, but overall mobility in post-secondary institutions as well as the

professional world has dramatically decreased as well, which may contribute to learning German as a second language losing its appeal.

Consequently, we are now facing a situation in which the contexts languages can actually be used in authentically have decreased as well. While the pandemic can be assumed to be a temporary problem, lower enrolment numbers and smaller course sizes have an influence on future course planning in institutional settings including employment decisions for language teachers.

For the reasons outlined above, research on difficulty should not only be approached comprehensively, by combining different approaches as was done in this study, it should also be seen as a cross-linguistic, dynamic concept that is very susceptible to changes in the learning environment and therefore requires constant attention from both research and practice.

References

A comparative analysis of relative difficulty (n.d.). Retrieved April 29, 2015, from https://www.nsa.gov/public_info/_files/cryptologic_spectrum/foreign_language.pdf

Aguado, K. (2012). Progression, Erwerbssequenzen und Chunks. Zur Lehr- und Lernbarkeit von Grammatik im Fremdsprachenunterricht. *AkDaF Rundbrief, 64*, 7-22.

Amman, U. (2018). *Die Stellung der deutschen Sprache in der Welt.* Berlin/Boston, MA: Mouton de Gruyter.

Ammar, A., & Lightbown, P. M. (2005). Teaching marked linguistic structures – More about the acquisition of relative clauses by Arab learners of English. In A. Housen & M. Pierrard (Eds.), *Investigation in instructed second language acquisition* (pp. 167-198). New York, NY: Mouton de Gruyter.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA/London: Harvard University Press.

Andrews, S. (1998). All these like little name things: A comparative study of language teachers' explicit knowledge of grammar and grammatical terminology. *Language Awareness, 8*, 143-159.

Andringa, S., & Rebuschat, P. (2015). New directions in the study of implicit and explicit learning. *Studies in Second Language Acquisition, 37*, 185-196.

Anton, C., Barske, T., Grabowski, J. & McKinstry, M. (2014). *Sag Mal. An Introduction to German Language and Culture*. Boston: Vista Higher Learning.

Arnawa, N. (2017). The implementation of natural semantic metalanguage and semantic field in language teaching: A case study. *Journal of Language Teaching and Research, 8*, 507-515.

Auswärtiges Amt (2015). *Deutsch als Fremdsprache weltweit. Datenerhebung 2015*. Retrieved April 26, 2015, from http://www.kmk.org/fileadmin/pdf/PresseUndAktuelles/2015/Bro_Deutschlernerhebung_final.pdf.

Ballestracci, S. (2005). *Zum Daf-Erwerb ausgewählter grammatischer Strukturen der deutschen Sprache bei italophonen Studierenden der Pisaner Fakultät der Lingue e Letteratura Straniere*. Pisa: Department for Linguistics.

BAMF, Bundesamt für Migration und Flüchtlinge (2020). *Aktuelle Zahlen. Tabellen, Diagramme, Erläuterungen*. Retrieved May 9, 2020, from https://www.bamf.de/SharedDocs/Anlagen/DE/Statistik/AsylinZahlen/aktuellezahlen-maerz-2020.html.

Bartley, D. (1970). The importance of attitude factors in language dropout: A preliminary investigation of group and sex differences. *Foreign Language Annals, 3*, 383-393.

Baten, K. (2011). Processability theory and German case acquisition. *Language Learning, 61*, 1-51.

Baten, K., & Willems, K. (2012). Case acquisition in the prepositional phrase from the standpoint of processability theory. *Deutsche Sprache, 40*, 221-239.

Bell, R. (1983). *An introduction to applied linguistics*. London: Batsford.

Berent, G. (1985). Markedness consideration in the acquisition of conditional sentences. *Language Learning, 35*, 337-372.

Berry, R. (2010). *Terminology in English language teaching: Nature and use.* Bern: Peter Lang.

Berry, R. (2009a). EFL majors' knowledge of metalinguistic terminology a comparative study. *Language Awareness, 18*, 113-128.

Berry, R. (2009b). Making the most of metalanguage. *Language Awareness*, *14*, 3-20.

Berry, R. (2004). Awareness of metalanguage. *Language Awareness*, *13*, 1-6.

Berry, R. (1997). Teachers' awareness of learners' knowledge: The case of metalinguistic terminology. *Language Awareness*, *6*, 136-146.

Bialystok, E. (1978) A theoretical model of second language learning. *Language Learning, 28*, 69-83.

Brandl, K. (2008). *Communicative language teaching in action: Putting principles to work.* Upper Saddle River, NJ: Pearson Prentice Hall.

Byrd, P. (2001). Textbooks: Evaluation for selection and analysis for implementation. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (3rd ed., pp. 415-442). Boston, MA: Heinle & Heinle.

Byrd, P. (1995). Issues in the writing and publication of grammar textbooks. In P. Byrd (Ed.), *Material writer's guide* (pp. 3-12). New York, NY: Heinle & Heinle.

Cajkler, W., & Hislam, J. (2002). Trainee teacher's grammatical knowledge: The tension between public expectation and individual competence. *Language Awareness*, *11*, 161-177.

Canadian Association of University Teachers of German (2013). *Enrollment report 2012/13*. Retrieved April 23, 2015 from http://www.cautg.org/wp-content/CAUTG-EnrolmentReport_Part1.pdf.

Chaplin, C., & Willard, N. (1936). *Modern times* [Video]. Youtube. https://www.youtube.com/watch?v=idG-dMSO2Qc.

Chavez, M. (2017). Hard rules and bad memories: College learners' accounts of what makes learning German grammar difficult. *Die Unterrichtspraxis, 50*(1), 1-21, 100.

Chavez, M. (2016). Students' accounts of grammatical forms of German that are difficult, unattainable, and irrelevant for self-expression. *Language Awareness, 25*(3), 197-221.

Chavez, M. (2013). How college learners of German envision their final attainment: Between community and future self. *Die Unterrichtspraxis, 46*, 75-90.

Chavez, M. (2007). Students' and teachers' assessments of the need for accuracy in the oral production of German as a foreign language. *The Modern Language Journal, 93*, 336-353.

Clahsen, H., Meisel, J., & Pienemann, M. (1983). *Deutsch als Zweitsprache: Der Spracherwerb auslaendischer Arbeiter*. Tübingen: Narr.

Clément, R., Smythe, P. C., & Gardner, R. C. (1978). Persistence in second-language study: motivational considerations. *Canadian Modern Language Review, 34*, 688-690.

Collins, L., Trofimovich, P., White, J., Cardoso, W., & Horst, M. (2009). Some input on the easy/difficult grammar question: An empirical study. *Modern Language Journal, 93*, 336-353.

Crabbe, D. (2003). The quality of language learning opportunities. *TESOL Quarterly, 37*(1), 9-35.

Cummins, J. (1991). Interdependence of first- and second-language proficiency in bilingual children. In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 70-89). Cambridge, England: Cambridge University Press.

Dahl, O. (2004). *The growth and maintenance of linguistic complexity.* Amsterdam: John Benjamins.

DeKeyser, R. M. (2016). Of moving targets and chameleons. Why the concept of difficulty is so hard to pin down. *Studies in Second Language Acquisition, 38*, 353-363.

DeKeyser, R. M. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 97-113). Mahwah, NJ: Lawrence Erlbaum Associates.

DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning, 55*, 1-25.

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 313-348). Oxford, England: Blackwell.

DeKeyser, R. M., & Sokalski, K. (1996). The differential role of comprehension and production practice. *Language Learning, 46*, 613-642.

Denhovska, N., Serratrice, L., & Payne, J. (2015). Acquisition of second language grammar under incidental learning conditions: The role of frequency and working memory. *Language Learning, 66*(1), 1-32.

Di Donate, R. & Clyde, M. (2015). *Deutsch Na Klar!: An Introductory German Course*. New York: McGrawHill.

Diehl, E., Pistorius, H., & Dietl, A. (2002). Grammatikerwerb im Fremdsprachenunterricht – Ein Widerspruch? In W. Borner & K. Vogel (Eds.), *Grammatik und Fremdsprachenerwerb* (pp. 143-163). Tübingen: Narr.

Diehl, E., Christen, H., Leuenberger, S., Pelvat, I., & Studer, T. (2000). *Grammatikunterricht: Alles fuer der Katz? Untersuchungen zum Zweitspracherwerb Deutsch.* Tübingen: Niemeyer.

Diewald, G., & Kresić, M. (2010). Ein übereinzelsprachliches kontrastives Beschreibungsmodell für Partikelbedeutungen. *Linguistik Online*, *44*(4), 5-18.

Dimroth, C., & Haberzettl, S. (2007). Je älter desto besser. Lernen Kinder Deutsch als Zweitsprache schneller als Deutsch als Erstsprache? In B. Ahrenholz, U. Bredel, W. Klein, M. Rost-Roth & R. Skiba (Eds.), *Empirische Forschung und Theoriebildung. Beiträge aus der Soziolinguistik, Gesprochene Sprache Forschung und Zweitspracherwerbsforschung* (pp. 227-239). Frankfurt am Main: Peter Lang.

Dirven, R. (1990). Pedagogical grammar. *Language Teaching, 23*(1), 1-18.

Dodeen, H. (2008). Assessing test-taking strategies of university students: developing a scale and estimating its psychometric indices. *Assessment and Evaluation in Higher Education, 33:4*, 409-419.

Dörnyei, Z. (2009). *The psychology of second language acquisition*. Oxford, England: Oxford University Press.

Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 114-138). Cambridge, England: Cambridge University Press.

Dubravac, V., & Pavicic Takac, V. (2013). The relationship between structure complexity and foreign language linguistic knowledge. *Jezikoslovlje, 14*, 577-598.

Dulay, H., & Burt, M. (1973). Should we teach children syntax? *Language Learning*, *23*, 245-258.

Dulay, H., Burt, M., & Krashen, S. (1982). *Language two.* New York, NY: Oxford University Press.

Elder, C., & Manwaring, D. (2004). The relationship between metalinguistic knowledge and learning outcomes among undergraduate students of Chinese. *Language Awareness*, *13*, 145-162.

Ellis, R. (2009). Measuring implicit and explicit knowledge of a second language. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philip & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3-25). Bristol, England: Multilingual Matters.

Ellis, R. (2008). Investigating grammatical difficulty in second language learning: Implications for second language acquisition research and language testing. *International Journal of Applied Linguistics*, *18*, 4-22.

Ellis, R. (2006). Modelling learning difficulty and second language proficiency: The differential contributions of implicit and explicit knowledge. *Applied Linguistics*, *27*(3), 431-463.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies on Second Language Acquisition, 27*, 141-172.

Ellis, R. (2001). Some thoughts on testing grammar: An SLA perspective. In C. Elder, A. Brown, E. Grove, K. Hill, T. Iwashita, T. Lumley, T. McNamara & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 251-263). Cambridge, England: Cambridge University Press.

Ellis, R. (1989). Are classroom and naturalistic acquisition the same? A study of the classroom acquisition of German word order rules. *SSLA, 11*, 305-328.

Ellis, R., & Loewen, S. (2007). Confirming the operational definitions of explicit and implicit knowledge in Ellis (2005): Responding to Isemonger. *Studies in Second Language Acquisition, 29*, 119-126.

Engel, U. (1987). Termini für Deutsch als Fremdsprache. In J. Ballweg, I. Klein, H. Steger & R. Wimmer (Eds.), *Sprache der Gegenwart* (pp. 192-214). Düsseldorf: Schwann.

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics, 27*, 423-449.

Felix, S. (1981). The effect of formal instruction on second language acquisition. *Language Learning,* 31, 87-112.

Fell, M. (2019, February 27). *Language learning: German and French drop by half in UK schools*. BBC. Retrieved May 9, 2020, from https://www.bbc.com/news/education-47334374.

Frieske, B., Öz, F., & Lorenz, L. (2004). *Die schwachen Verben*. [Seminararbeit, Universität Bielefeld]. Retrieved from https://www.uni-bielefeld.de/lili/personen/useelbach/STUD/studis.html?__xsl=/unitemplate_2009_print.xsl.

Funk, H., Bayerlein, O., Christiany, C., Demme, S., Kuhn, C. & Winzer-Kiontke, B. (2006) *Studio D Deutsch als Fremdsprache Grundstufe*. Gesamtband. Berlin: Cornelsen.

Funk, H. (1995). Grammatikvermittlung in Deutsch-als-Fremdsprache-Lehrwerken: Historische und aktuelle Perspektiven. In C. Gnautzmann & F. G. Königs (Eds.), *Perspektiven des Grammatikunterrichts* (pp. 29-46). Tübingen: Narr.

Gallagher-Brett, A. (2004). *Seven hundred reasons for studying languages*. The Higher Education Academy. Retrieved August 1, 2021, from https://eprints.soton.ac.uk/362669/1/700_reasons.pdf

Gánem-Gutiérrez, G., & Roehr, K. (2013). Use of L1, metalanguage, and discourse markers: L2 learners' regulation during individual task performance. *International Journal of Applied Linguistics*, *21*(3), 297-318.

Gardner, R. C. (2010). *Motivation and second language acquisition: The socio-educational model.* New York, NY: Peter Lang.

Gardner, R. C., & Lambert, W. E. (1959). Motivational variables in second-language acquisition. *Canadian Journal of Psychology, 13*, 266-272.

Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to fluency. *The Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes, 61*(3), 325-353.

Gebhard, M., Chen, I., & Britton, L. (2014). Miss, nominalization is a nominalization: English language learners' use of SFL metalanguage and their literacy practices. *Linguistics and Education, 26*, 106-125.

Gongolweski, M, Moser, B. & Partsch, C. (2013). *Treffpunkt Deutsch Grundstufe*. Boston: Pearson.

Gouin, F. (1892). *The art of teaching and studying languages*. London: George Philip & Son.

Graddol, D. (1997). *The future of English.* London: British Council.

Graus, J., & Coppen, P. (2016). Student teacher beliefs on grammar instruction. *Language Teaching Research, 20*(5), 571-599.

Graus, J., & Coppen, P. (2015). Defining grammatical difficulty: A student teacher perspective. *Language Awareness, 24*(2), 101.

Gregg, K. R. (1984). Krashen's monitor and occam's razor. *Applied Linguistics, 5*(2), 79-100.

Grimm, J. (1819). *Deutsche Grammatik. Erster Theil*. Göttingen: Dieterichsche Buchhandlung.

Grün Johansson, E. (2018). *Die Rolle der expliziten Grammatikvermittlung beim Lernen von Deutsch als Fremdsprache: Eine kontrastive Lehrbuchanalyse von drei Lehrbüchern der Stufe 1 des schwedischen Gymnasiums*. [Dissertation, Linnaeus University]. Digitala Vetenskapliga Arkivet.

Guerrettaz, A. M., & Johnston, B. (2013). Materials in the classroom ecology. *The Modern Language Journal, 97*, 779-796.

Gutiérrez, X. (2013a). The construct validity of grammaticality judgement tests as measures of implicit and explicit knowledge. *Studies in Second Language Acquisition, 35*(3), 423-449.

Gutiérrez, X. (2013b). Metalinguistic knowledge, metalingual knowledge, and proficiency in L2 Spanish. *Language Awareness*, *22*, 176-191.

Gutiérrez, X. (2012). Implicit knowledge, explicit knowledge, and achievement in second language(L2) Spanish. *Canadian Journal of Applied Linguistics*, *15*, 20-41.

Habersack, C., Pude, A. & Specht, F. (2012). *Menschen A1/A2. Deutsch als Fremdsprache*. München: Hueber.

Haberzettl, S. (2006). Progression im ungesteuerten Erwerb und im gesteuerten Erwerb. In B. Ahrenholz (Ed.), *Kinder mit Migrationshintergrund. Spracherwerb und Fördermöglichkeiten* (pp. 203-220). Freiburg: Fillibach.

Haberzettl, S. (2005). *Der Erwerb der Verbstellungsregeln in der Zweitsprache Deutsch durch Kinder mit russischer und türkischer Muttersprache*. Tübingen: Niemeyer.

Haeuser, R. (2012, May 25). *Schüler in Europa: Keine Lust auf Deutsch*. Deutsche Welle. Retrieved April 29, 2015, from http://www.dw.de/sch%C3%BCler-in-europakeine-lust-auf-deutsch/a-15977776.

Hawkins, J. (2015). *A comparative typology of English and German: Unifying the contrasts.* London: Croom Helm.

Heringer, H. J., & Keller-Bauer, F. (1984). Probleme einer gebrauchsgrammatischen Terminologie. *Sprache und Literatur in Wissenschaft und Unterricht, 53*, 65-86.

Holliday, A. (2006). Native-speakerism. *ELT Journal*, *60*, 385-387.

Holt, R. F. (2006). Persistence factors in secondary school additional language study. *Journal of Language and Learning, 5*, 86-97.

Housen, A. (2014). Difficulty and complexity of language features and second language. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (Vol. 4, pp. 2205-2213). Hoboken, NJ: John Wiley & Sons.

Housen, A., & Simoens, H. (2016). Introduction: Cognitive perspectives on difficulty and complexity in L2 acquisition. *Studies in Second Language Acquisition, 38*(2), 163-175.

Housen, A., Pierrard, M., & Van Daele, S. (2005). Rule complexity and the effectiveness of explicit grammar instruction. In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language acquisition* (pp. 235-270). Berlin: Mouton de Gruyter.

Hornby, A.S. (2000). *Oxford Advanced Learner's Dictionary*. Oxford: Oxford University Press.

Hsieh, P. (2005). *How college students explain their grades in a foreign language course: The interrelationship of attributions, self-efficacy, language learning beliefs, and achievement.* Ann Arbor, MI: ProQuest LLC.

Hu, G. (2011). Metalinguistic knowledge, metalanguage, and their relationship in L2 learners. *System*, *39*(1), 63-77.

Hu, G. (2002). Psychological constraints on the utility of metalinguistic knowledge in second language production. *Studies in Second Language Acquisition*, *24*, 347-386.

Hulstijn, J. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning. *Studies in Second Language Acquisition, 27*, 129-140.

Ivančić, B. (2010). Grammatische Terminologie im Kontrast. Einige Überlegungen aus der Sicht des DaF-Unterrichts in Italien. *Linguistik online, 41*(1), 41-50.

Izumi, Y., & Izumi, S. (2004). Investigating the effects of oral output on the learning of relative clauses in English: Issues in the psycholinguist requirements for effective output tasks. *The Canadian Modern Language Review, 60*, 587-609.

James, C., & Garret, P (1992). The scope of awareness. In C. James & P. Garrett (Eds.), *Language awareness in the classroom* (pp. 3-20). London: Longman.

Jansen, L. (2008). Acquisition of German word order in tutored learners: A cross-sectional study in a wider theoretical context. *Language Learning, 58*(1), 185-231.

Johnson, K., & Johnson, H. (1998). *Encyclopedic dictionary of applied linguistics: A handbook for language teaching.* Oxford, England: Blackwell Reference Online.

Joint Council for Qualification (2018). *A review of modern foreign languages at A level: A* grade and low take up*. Retrieved May 6, 2020, from: https://www.jcq.org.uk/media-centre/newsreleases/mfl-review-press-notice.

Kang, E. Y., Sok, S., & Han, Z. (2019). Thirty-five years of ISLA on form-focused instruction: A meta-analysis. *Language Teaching Research, 23*(4), 403-427.

Keen, J. (1997). Grammar, metalanguage and writing development. *Teacher Development, 1*(3), 431-445.

Kleineidam, H. (1986). *Fremdsprachengrammatik: Analysen und Positionen. Beiträge mit dem Schwerpunkt Französisch*. Tübingen: Narr.

Krahnke, K. (1987). *Approaches to syllabus design for foreign language learning*. Englewood Cliff, NJ: Prentice Hall Regents.

Krashen, S. D. (1981). *Second language acquisition and second language learning.* London: Pergamon.

Krashen, S. D. (1978). Individual variation in the use of the monitor. In W. Ritchie (Ed.) *Second language acquisition research* (pp. 175-183). New York, NY: Academic Press.

Krashen, S. D., & Seliger, H. (1975). Maturational constraints in the acquisition of a second language and second dialect. *Language Sciences, 38,* 28-29.

Kwaakernaak, J. (2007). Babylonia Grammatikprogression und der GER. *Babylonia, 1*, 34-52.

Lamendella, J. (1977). General principles of neurofunctional organization and their manifestations in primary and non-primary language acquisition. *Language Learning, 27*, 155-196.

Lemke, C., Rohrmann, L. & Scherlin, T. (2009). *Berliner Platz 1 Neu. German for Beginners*. Berlin: Langenscheidt.

Lincoln, Y., & Guba, E. G. (1985): *Naturalistic inquiry*. Beverly Hills, CA: Sage.

Lindstrom, E. (2008). *Language complexity and interlinguistic difficulty. Language complexity: Typology, contact, change.* Amsterdam: John Benjamins.

Macken-Horarik, M. (2009). Navigational metalanguages for new territory in English: The potential of grammatics. *English Teaching: Practice and Critique*, *8*(3), 55-69.

Macken-Horarik, M. (2008). A “good enough” grammatics: Developing an effective metalanguage for school English in an era of multiliteracies. In C. Wu, C. Matthiessen & M. Herke (Eds.), *Proceedings of the ISFC 35: Voices around the world* (pp. 43-48). Sydney: 35th ISFC Organizing Committee.

Mackey, W. F. (1965). *Language Teaching Analysis*. London: Longman.

Maijala, M. (2010). Grammatische Unterrichtsinhalte in Lehrwerken Deutsch als Fremdsprache. *Zielsprache Deutsch, 3,* 17-44.

McLaughlin, B. (1978). The monitor model: Some methodological considerations. *Language Learning, 28*, 309-332.

Meyer, S. (2013). *Foreign language enrolment attribution: Exploring the key factors of motivation, academic success, instruction, and anxiety at the elementary level.* Ann Arbor, MI: ProQuest LLC.

Modern Language Association (2016). *Enrollments in languages other than English in United States institutions of higher education*. Retrieved April 2, 2021 from https://www.mla.org/Resources/Research/Surveys-Reports-and-OtherDocuments/Teaching-Enrollments-and-Programs/Enrollments-in-Languages-OtherThan-English-in-United-States-Institutions-of-Higher-Education.

Mueller, T., & Harris, R. (1966). The effect of an audio-lingual program on drop-out rate. *The Modern Language Journal, 50*, 133-137.

Murakami, A. (2013, August). *A dynamic approach to the longitudinal development of L2 English grammatical morphemes* [Conference presentation]. EUROSLA 23, University of Amsterdam, Amsterdam, the Netherlands.

Neuner, G. (1996). Progressionsverfahren bei Lehrwerkentwicklung. Beispiel: „Grammatik“. In H. Popp (Ed.), *Deutsch als Fremdsprache. An den Quellen eines Fachs* (pp. 685-706). München: Iudicisum Verlag.

Newby, D. (2001). *Pedagogical grammar*. Retrieved September 2, 2018, from https://static.uni-graz.at/fileadmin/gewi-institute/Anglistik/Fachdidaktik/Downloads/angl3www_newby_encyclopedia.pedgramm.pdf.

Noels, K. A., Clément, R., & Pelletier, L. G. (1999). Perceptions of teachers’ communicative style and students’ intrinsic and extrinsic motivation. *The Modern Language Journal, 83*, 23-34.

Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analyses. *Language learning, 50*, 417-528.

Ortega, L. (2014). Ways forward for bi/multlingual turn in SLA. In S. May (Ed.), The multilingual turn: Implications for SLA, TESOL, and bilingual education (pp.32-53). New York: Routledge/Taylor & Francis.

Palmer, H.E. (1917). *The scientific study and teaching of languages*. London: Harrap & Company.

Palotti, G. (2015). A simple view of linguistic complexity. *Second Language Research, 31*, 117-134.

Penning-Hiemstra, S., Sprech, F., Wagner, D. & Bovermann, M. (2017). *Schritte A1/A2. München*: Hueber.

Perkins, K., & Freeman, D. L. (1975). The effect of formal language instruction on the order of morpheme acquisition. *Language Learning, 25*(2), 237-243.

Phipps, S. (2015). *Concerns in Scotland over modern foreign languages*. Retrieved April, 29, 2015, from http://www.sec-ed.co.uk/news/concerns-inscotland-over-modern-foreign-languages.

Pica, T. (1983). Adult acquisition of English as a second language under different conditions of exposure. *Language Learning, 33*(4), 465-497.

Pienemann, M. (2007). Processability theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 137-154). Mahwah, NJ: Erlbaum.

Pienemann, M. (1998). Developmental dynamics in L1 and L2 acquisition: Processability theory and generative entrenchment. *Bilingualism, 1*(1), 1-20.

Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics, 10*, 52-79.

Pienemann, M. (1985). Learnability and syllabus construction. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and assessing second language development* (pp. 23-76). Clevedon, England: Multilingual Matters.

Pienemann, M. (1984). Psychological constraints on the teachability of languages. *Studies in Second Language Acquisition, 6*(2), 186-214.

Pienemann, M., & Hakansson, G. (1999). A unified approach towards the development of Swedish as L2: A processability account. *Studies in Second Language Acquisition, 21*(3), 383-420.

Pienemann, M., & Johnston, M. (1986). An acquisition based procedure for language assessment. *Australian Journal of Applied Linguistics, 9*, 92-122.

Pienemann, M., & Kessler, J. U. (Eds.). (2011). *Studying processability theory: An introductory textbook.* Amsterdam: John Benjamins Publishing.

Pratt, C., Agnello, M. F., & Santos, S. (2009). Factors that motivate high-school students' decisions to study Spanish. *Hispania, 92*, 800-813.

Rahimpour, M. (2010). Current trends on syllabus design in foreign language instruction. *Procedia – Social and Behavioral Sciences, 2*(2), 1660-1664.

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning, 63*(3), 595-626.

Reid, J. (1995). Developing ESL writing materials for publication or writing as a learning experience. In P. Byrd (Ed.), *Material writer's guide* (pp. 64-78). New York, NY: Heinle & Heinle.

Renou, J. (2001). An examination of the relationship between metalinguistic awareness and second-language proficiency of adult learners of French. *Language Awareness*, *10*, 248-267.

Ritterbusch, R. (2006). Learner difficulties with German case: Implications from an action research study. *Die Unterrichtspraxis*, *39*, 30-45.

Robinson, P. (2001). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instructions* (pp. 287-318). Cambridge, England: Cambridge University Press.

Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search, and instructed conditions. *Studies in Second Language Acquisition, 18*(1), 27-67. doi:10.1017/S0272263100014674

Rodgers, T. (1971). The concept of linguistic difficulty. *University of Hawaii Working Papers in Linguistics, 3*, 109-120.

Rodrigues, E. N. (2015). *Curriculum design and language learning: An analysis of English textbooks in Brazil*. [Dissertation, Andrews University]. Retrieved from https://digitalcommons.andrews.edu/cgi/viewcontent.cgi?article=1000&context=dissertations.

Rodríguez Silva, L. H. , & Roehr-Brackin, K. (2016). Perceived learning difficulty and actual performance: Explicit and implicit knowledge. *Studies in Second Language Acquisition, 38*(2), 317-340.

Roehr, K. (2008). Metalinguistic knowledge and language ability in university-level L2 learners. *Applied Linguistics, 29*, 173-199.

Roehr, K. (2006). Metalinguistic knowledge in L2 task performance: A verbal protocol analysis. *Language Awareness, 15*, 180-198.

Roehr, K., & Gáneem-Gutiérrez, G. (2009). The status of metalinguistic knowledge in instructed adult L2 learning. *Language Awareness*, *18*, 165-181.

Rogers, M. (1987). Learners´ difficulties with grammatical gender in German as a foreign language. *Applied Linguistics*, *8*(1), 48-66.

Saussure, F.d. (1976). *Grundfragen der allgemeinen Sprachwissenschaft*. Berlin: Mouton de Gruyter.

Sevin, D. & Sevin, I. (2007). *Wie geht's? An Introductory German Course*. Boston: Heinle.

Scheffler, P. (2009). Rule difficulty and the usefulness of instruction. *ELT Journal, 63*, 5-12.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129-158.

Seliger, H. W. (1979). On the nature and function of language rules in language teaching. *TESOL Quarterly, 13*, 359-369.

Serrano, R. (2001). From metalinguistic instruction to metalinguistic knowledge, and from metalinguistic knowledge to performance in error correction and oral production tasks. *Language Awareness*, *20*, 1-16.

Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition, 15*, 165-179.

Shiu, L. J. (2011). *EFL learners' perceptions of grammatical difficulty in relation to second language proficiency, performance and knowledge*. Ottawa, ON: Library and Archieves Canada.

Sick, B. (2016). *Der Dativ ist dem Genitiv sein Tod – Folge 4-6*. Köln: Kiepenheuer & Witsch.

Sick, B. (2004). *Der Dativ ist dem Genitiv sein Tod. Ein Wegweiser durch den Irrgarten der deutschen Sprache*. Köln: Kiepenheuer & Witsch.

Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition, 15*(2), 165-179.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning, 60*, 263-308.

Spada, N., Lightbown, P. M., & White, J. (2005). The importance of meaning in explicit form focused instruction. In A. Housen, & A. Pierrard (Eds.), *Current issues in instructed second language learning* (pp. 199-234). Brussels: Mouton de Gruyter.

Stern, H.H. (1992). *Issues and options in language teaching*. Oxford University Press.

Stern, H. H. (1983). *Fundamental concepts of language teaching*. Oxford, England: Oxford University Press.

Suzuki, Y., & DeKeyser, R. M. (2017). The interface of explicit and implicit knowledge a second language: Insights from individual differences in cognitive aptitudes. *Language Learning, 67*(4), 747-790.

Suzuki, Y., & DeKeyser, R. M. (2015). Comparing elicited imitation and word monitoring as measures of implicit knowledge. *Language Learning, 65*(4), 860-895.

Tammenga-Helmantel, M. (2012). Zur Bestimmung des Stellenwerts von Grammatik in DaF-Lehrwerken – Am Beispiel von Schulbüchern in der niederländischen Unterstufe. *Deutsch als Fremdsprache, 49*, 67-77.

Tinsley, T., & Board, K. (2013). *Language learning in primary and secondary schools in England: Findings from the 2012 Language Trends survey.* Reading: CfBT Education Trust.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research, 45*, 89-125.

Tomlinson, B. (2012). Materials development for foreign language learning and teaching. *Language Teaching, 45*, 143-179.

Tschirner, E., Nikolai, B.& Terrell, T.D. (2012). *Kontakte. A Communicative Approach.* New York: McGrawHill.

Tschirner, E. (1999). Lernergrammatiken und Grammatikprogression. In B. Skibitzki & B. Wotjqak (Eds.), *Linguistik und Deutsch als Fremdsprache. Festschrift für Gerhard Helbig* (pp. 227-240). Tübingen: Niemeyer.

Tudor, I. (1992). Learner-centredness in language teaching: Finding the right balance. *System*, *20*(1), 31-44.

Twain, M. (1880). *The awful German language*. Retrieved September 14, 2018, from https://www.cs.utah.edu/~gback/awfgrmlg.html.

Vafaee, P., Suzuki, Y., & Kaschinske, I. (2017). Validating grammaticality judgement tests. Evidence from two new psycholinguistic measures. *Studies in Second Language Acquisition, 39*(1), 59-95.

Vainikka, A., & Young-Scholten, M. (1996). *Studies on language acquisition: Acquisition of German: Introducing organic grammar.* Berlin: Mouton de Gruyter.

Van Moere, A. (2012). A psycholinguistic approach to oral language assessment. *Language Testing, 29*, 325-344.

VanPatten, B. (1993). Grammar teaching for the acquisition-rich classroom. *Foreign Language Annals, 26*, 435-450.

Wesely, P. M. (2012). Learner attitudes, perceptions, and beliefs in language learning. *Foreign Language Annals, 45*, 98-117.

Wesely, P. M. (2010): Student attrition from traditional and immersion foreign language programs. *Language and Linguistic Compass, 4*, 804-817.

Westhoff, G. (2007). Grammatische Regelkenntnisse und der GER. *Babylonia, 1*, 12-21.

White, J., & Ranta, L. (2002). Examining the interface between metalinguistic task performance and oral production in second language. *Language Awareness, 11*(4), 259-290.

Widdowson, H. G. (1997). Metalanguage and interlanguage. In R. Hickey & S. Puppel (Eds.), *Language history and linguistic modelling* (pp. 1887-1899). Berlin: Mouton de Gruyter.

Wilkins, D. A. (1976). *Notional syllabuses*. London: Oxford University Press.

Winkler, S. (2011). Progressionsfolgen im DaF-Unterricht. Eine Interventionsstudie zur Vermittlung der deutschen (S)OV-Wortstellung. In N. Hahn & T. Roelcke (Eds.), *Materialien Deutsch als Fremdsprache* (Vol. 85, pp. 193-207). Göttingen: Universitätsverlag Göttingen.

Appendix

Appendix A: Learners' Perceptions of Grammatical Difficulty Questionnaire

Learners' perception of grammatical difficulty

The following is a list of German grammatical features. Please indicate on a scale from 1-6, whether a particular grammatical feature has been more or less difficult by circling only **one number**. Please be sure to respond to all features that you have learned so far. If you have not been taught a specific feature, please circle *Not been taught yet.* Each feature comes with examples. Please do not **make a decision based on the difficulty of the example** – they only serve to demonstrate what a certain feature means.

Note **1 = not at all difficult** **6 = extremely difficult**

	Past tense							
1	**Choice of correct auxiliary** *Ich habe/bin gestern eingekauft.*	1	2	3	4	5	6	Not been taught yet
	Choice of correct participle							
2	*Ich bin geschwommen/geschwimmt/schimmen*	1	2	3	4	5	6	Not been taught yet
	Word order							
3	**Position of verb** (dep, independent clauses, questions etc.) *Ich lese jeden Tag/Ich jeden Tag lese.* *Wenn ich traurig bin, gehe ich ins Kino*/Wenn ich bin traurig, ich gehe ins Kino.	1	2	3	4	5	6	Not been taught yet
4	**Adverb placement (time and place)** *Ich gehe um 20 Uhr ins Bett/Ich gehe ins Bett um 20 Uhr.*	1	2	3	4	5	6	Not been taught yet
	Prepositions							
5	**Deciding which one to use** *Ich gehe in/zu die Uni.*	1	2	3	4	5	6	Not been taught yet
	Two-way prepositions							
6	**Choosing the correct preposition** *Ich hänge das Bild an/auf/neben die Wand.*	1	2	3	4	5	6	Not been taught yet
7	**Choosing the correct case to go with the prep.** *Ich lege die Katze auf das/dem Bett.*	1	2	3	4	5	6	Not been taught yet
8	**Separable prefix-verbs** *Ich stehe um 6 Uhr auf/Ich aufsteh um 6 Uhr.*	1	2	3	4	5	6	Not been taught yet

9	**Verb endings (subject/ verb agreement)** *Ich geh**e**/geh**t** in die Bibliothek.*	1	2	3	4	5	6	Not been taught yet
10	**Gern-construction** *Ich gehe gern ins Kino/Ich gern Kino.*	1	2	3	4	5	6	Not been taught yet
11	**Negation (kein/keine versus nicht)** *Peter geht nicht/geht in keine Bäckerei.*	1	2	3	4	5	6	Not been taught yet
12	**Artikel** **Choice of definite versus indefinite article** *Ein Mann versus der Mann hat kurzes Haar.*	1	2	3	4	5	` 6	Not been taught yet
13	**Gender (masculine, feminine,neuter)** *Der/die/das Kind spielt gern im Garten.*	1	2	3	4	5	6	Not been taught yet
14	**Accusative case** *Ich habe ein Bruder/einen Bruder.*	1	2	3	4	5	6	Not been taught yet
15	**Dative case** *Ich gebe meinem Bruder/mein Bruder ein Buch.*	1	2	3	4	5	6	Not been taught yet
16	**Possessive Pronouns** **Use of correct pronoun** *Meine/Deine/Ihre/Mutter ist hübsch***.**	1	2	3	4	5	6	Not been taught yet
17	**Use of correct ending** *Mein**e**/Mein Freundin hat rote Haare.*	1	2	3	4	5	6	Not been taught yet
18	**Modal verbs** **Placement** *Ich muss heute Abend mein Zimmer putzen/Ich heute Abend mein Zimmer putzen muss.*	1	2	3	4	5	6	Not been taught yet
19	**Use of correct modal verb** *Ich kann/muss/soll meine Hausaufgaben machen.*	1	2	3	4	5	6	Not been taught yet
20	**Article endings** **Figuring out which ending to use** *Ich habe einen roten/rote/rotes Auto*	1	2	3	4	5	6	Not been taught yet

In the following section, you are asked to **explain why** a specific feature (only those that are numbered, not the highlighted category titles) is not difficult at all (1 or 2), which one is moderately difficult (3 and 4) and which one is extremely difficult (5 and 6). Please make sure that you name three **different features** instead of three aspects of only one feature. Please answer in **as much detail as possible** and feel free to elaborate on different contexts such as using, studying, understanding etc. You are also welcome to provide **examples or anecdotes** of your learning experience in addition to mere explanations.

1. Please indicate which of the above mentioned feature was **not difficult at all (1 or 2 on the scale)** and explain why. Please explain in detail.

I ranked feature ____ (only one) as **not difficult at all** because:

_____

_____

_____

_____

_____

_____

_____

_____

2. Please indicate which of the above mentioned feature was **moderately difficult (3 or 4 on the scale)** and explain why. Please explain in detail.

I ranked feature ____ (only one) as **moderately difficult** because:

_____

_____

_____

_____

_____

_____

_____

_____

3. Please indicate which of the above mentioned feature was **extremely difficult (5 or 6 on the scale)** and explain why. Please explain in detail.

I ranked feature ____ (only one) as **extremely difficult** because:

_____

_____

_____

_____

_____

_____

_____

_____

Additional comments

Here, you have the opportunity to leave additional comments. Is the chart missing something essential? Would you like to elaborate further on one of the aspects or features that are mentioned? Do you have additional thoughts about the concept of *difficulty* that have not been covered so far?

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Appendix B: Sentences and Instructions for Oral Imitation Test

Use of experiment software Open Sesame – delivered on computers:

Screen 1: instructions

“You will hear a German sentence that contains a statement. Then, you will be asked to agree (click XY) or disagree (click XM). Finally, you will be asked to repeat the sentence in correct German. Click ZZ to record.”

Screen 2: audio input (see sentences below)

Screen 3: Agree/Disagree (click)

Screen 4: Students repeat / reconstruct sentence they heard (screen two) and it will be recorded

Sentences:

- Am Wochenende habe ich ins City Center gegangen.
- Als Kind bin ich oft nach Deutschland gereist.
- Ich habe im Sommer sehr viel geschwimmt.
- Gestern Abend habe ich Fisch gegessen.
- Im Semester gehe ich ins Bett um 8 Uhr.
- Meine Familie spielt im Sommer im Garten Fussball.
- An den Wochenenden ich arbeite nicht sehr viel.
- Wenn ich traurig bin, esse ich Schokolade.
- Ich habe keine Freizeit, weil ich habe viele Kurse an der Uni.
- Jeden Tag komme ich um 8 Uhr an der Uni an.
- Am Freitag aufstehe ich um 5 Uhr am Morgen.
- Die Augen von meinem Vater ist gruen.
- Ich habe blonde Haare.
- Ich habe nicht eine grosse Familie.
- Meine Oma hat keine Schwester.
- Ich moechte heute Abend mein Zimmer putzen.
- Jeden Tag muss ich essen Fruehstueck am Morgen.
- In meinem Zimmer haengt ein grosses Foto von Edmonton.

- Mein Deutschprofessor hat blaues Haare.
- Ich schenke mein Mutter zum Geburtstag einen Kuchen.
- Ich gebe deinem Bruder ein Nutellatoast zum Fruehstueck.
- Ich habe in diesem Semester eine Mathekurs.
- Ich moechte nach der Universitaet einen Job finden.
- Nach der Uni finde ich einen Job.

Appendix C: Sentences and Instructions for Timed and Untimed Grammaticality Judgement Test

Timed: implicit knowledge – sentence is shown on ppt, students have to decide "correct / incorrect" / "don't know" / "not enough time" on a pen-and-paper form

Sentence 1: correct □ incorrect □ I don't know □ I didn't have enough time □
Sentence 2: correct □ incorrect □ I don't know □ I didn't have enough time □
Sentence 3: correct □ incorrect □ I don't know □ I didn't have enough time □

….

Untimed: explicit knowledge (delivered on a different day than the timed but it contains the same sentences) – pen-and-paper format

For both tests, sentences are randomized to avoid order effects; for each grammatical structure, the students see four sentences (2 correct, 2 incorrect)

Participant version of untimed GJT, please see below:

Untimed Grammaticality Judgement Test

Please read the sentences below and decide whether they are grammatically correct or incorrect. You do not need to correct the mistake (if the sentence is wrong), nor do you have to explain why. This test is exclusively about deciding whether the sentence is correct or incorrect.

Meine Schwester und ich sind gestern zum Abendessen Spaghetti gekocht.
correct □ incorrect □ I don't know □

Herr Koch hat jeden Morgen um 6 Uhr am Morgen aufgestanden.
correct □ incorrect □ I don't know □

In den Sommerferien haben wir sehr viel Sport gemacht.
correct □ incorrect □ I don't know □

Zu Weihnachten ist meine komplette Familie zu Besuch gekommen.
correct □ incorrect □ I don't know □

Mein Bruder ist gestern 20 km gelaufen.
correct □ incorrect □ I don't know □

Simone hat gestern vier Stunden lang mit Peter telefoniert.
correct □ incorrect □ I don't know □

Unser Deutschkurs hat gestern einen sehr langen Test geschreibt.
correct □ incorrect □ I don't know □

Alex und Sandra sind noch nie im Meer geschwimmt.
correct □ incorrect □ I don't know □

Wenn das Wetter ist warm, gehe ich ins Kino oder bleibe zuhause.
correct □ incorrect □ I don't know □

Ich lese ein Buch, weil ich Fernsehgucken nicht mag.
correct □ incorrect □ I don't know □

Weil ich bin krank, ich schlafe und trinke Ginger Ale.
correct □ incorrect □ I don't know □

Ich liege in der Sonne, wenn das Wetter warm ist.
correct □ incorrect □ I don't know □

In den Ferien Susanne ist mit ihren Freundinnen an den Strand gefahren.
correct □ incorrect □ I don't know □

Jeden Morgen ich trinke Kaffee und esse Toast.
correct □ incorrect □ I don't know □

Peters Französischkurs beginnt jeden Mittwoch um 19 Uhr.
correct □ incorrect □ I don't know □

Meine Wohnung hat ein großes Schlafzimmer und eine kleine Küche.
correct □ incorrect □ I don't know □

Frau Wagners Kinder spielen im Garten jedes Wochenende.
correct □ incorrect □ I don't know □

Die Universität Hamburg feiert auf dem Campus nächstes Jahr ein großes Fest.
correct □ incorrect □ I don't know □

Peter Smith faehrt jedes Wochenende zu IKEA.
correct □ incorrect □ I don't know □

Am Montag einkauft Susan immer bei Superstore.
correct □ incorrect □ I don't know □

Matthias und Konrad aufräumen nie ihre Wohnung.
correct □ incorrect □ I don't know □

Am liebsten sehen die Wagners am Abend fern.
correct □ incorrect □ I don't know □

Montags und dienstags stehe ich immer um 7 Uhr auf.
correct □ incorrect □ I don't know □

Monika und Larissa geht gerne schwimmen.
correct □ incorrect □ I don't know □

Frau Wagner hast zwei Brüdern, eine Schwester und viele Onkel und Tanten.
correct □ incorrect □ I don't know □

Tom und Timo sind seit vielen Jahren gute Freunde und wohnen auch zusammen in einer Wohnung.
correct □ incorrect □ I don't know □

Meine Nachbarn kochen am Wochenende den ganzen Tag Essen.
correct □ incorrect □ I don't know □

Ich habe nicht eine Schwester.
correct □ incorrect □ I don't know □

Daniel trinkt keinen Kaffee zum Frühstück.
correct □ incorrect □ I don't know □

Simone geht nicht gern ins Kino.
correct □ incorrect □ I don't know □

Herr Mueller hat heute nicht Kopfschmerzen, weil er lange geschlafen hat.
correct □ incorrect □ I don't know □

Meine Freundin kann sehr gut backen einen Schokoladenkuchen.
correct □ incorrect □ I don't know □

Sandra ist in Mathe sehr schlecht und sie muss lernen jeden Tag.
correct □ incorrect □ I don't know □

Der Arzt sagt, mein Vater soll jeden Morgen 3 Kilometer joggen, aber er ist zu faul.
correct □ incorrect □ I don't know □

Viele Menschen wollen schöne Autos und teure Häuser haben.
correct □ incorrect □ I don't know □

Meine Nachbarin hat ein schwarzes Katze.
correct □ incorrect □ I don't know □

Schnelles Autos sind oft sehr teuer und sie brauchen sehr viel Benzin.
correct □ incorrect □ I don't know □

Am liebsten schaue ich interessante Filme und Dokumentationen.
correct □ incorrect □ I don't know □

Meine Schwester isst gerne süße Bananen.
correct □ incorrect □ I don't know □

Ich schenke dem Bruder zum Geburtstag ein Buch.
correct □ incorrect □ I don't know □

Meine Freundin Jessica hatte einen Autounfall.
correct □ incorrect □ I don't know □

Von Januar bis April hat Frau Millers Tochter dem Deutschkurs mit Prof. Hohenstein.
correct □ incorrect □ I don't know □

In Deutschland hat Sophie ein Wohnung mit großem Garten und vielen Fenstern.
correct □ incorrect □ I don't know □

Appendix D: Metalinguistic Knowledge Test

Research Code: _____

Metalinguistic Knowledge Test

All sentences below contain **one mistake** each, which is underlined. This test **is not** about providing the correct answer, i.e. correcting the mistake but about **explaining** (in grammatical terms) **WHY** the underlined part is wrong. Please don't shy away from providing an explanation just because you don't feel you have the grammatical terms. In that case, use your own words! Please circle a,b,c or d.

1. ***Studenten an der University of Alberta <u>können kaufen</u> auf dem Campus viele Bücher.***
   a. I don't know why this part of the sentence is underlined, i.e. wrong.
   b. I know that the underlined part is wrong but I can't explain why.
   c. The underlined part is wrong because

      _____

      _____

   d. This feature hasn't been taught yet.

2. ***Im Sommer 2016 <u>haben</u> meine Freunde und ich jedes Wochenende in die Disko gegangen.***
   a. I don't know why this part of the sentence is underlined, i.e. wrong.
   b. I know that the underlined part is wrong but I can't explain why.
   c. The underlined part is wrong because

      _____

      _____

   d. This feature hasn't been taught yet.

3. ***Weil ich <u>habe</u> viele Kurse, kann ich meine Freunde nicht sehen.***
   a. I don't know why this part of the sentence is underlined, i.e. wrong.
   b. I know that the underlined part is wrong but I can't explain why.
   c. The underlined part is wrong because

      _____

      _____

   d. This feature hasn't been taught yet.

4. ***Ich habe <u>ein Bruder</u> und er hilft mir oft mit den Mathematikhausaufgaben.***
   a. I don't know why this part of the sentence is underlined, i.e. wrong.
   b. I know that the underlined part is wrong but I can't explain why.
   c. The underlined part is wrong because

      _____

      _____

   d. This feature hasn't been taught yet.
5. ***Meine Lehrerin kommt <u>in die Klasse jeden Morgen um 9 Uhr.</u>***
   a. I don't know why this part of the sentence is underlined, i.e. wrong.
   b. I know that the underlined part is wrong but I can't explain why.
   c. The underlined part is wrong because

      _____

      _____

   d. This feature hasn't been taught yet.

6. ***Viele Menschen <u>aufstehen</u> sehr früh, wenn sie arbeiten müssen.***
   a. I don't know why this part of the sentence is underlined, i.e. wrong.
   b. I know that the underlined part is wrong but I can't explain why.
   c. The underlined part is wrong because

      _____

      _____

   d. This feature hasn't been taught yet.

7. ***Sabine, Peter und Michael <u>fährt</u> am Nachmittag oft ins Fitnessstudio.***
   a. I don't know why this part of the sentence is underlined, i.e. wrong.
   b. I know that the underlined part is wrong but I can't explain why.
   c. The underlined part is wrong because

      _____

      _____

   d. This feature hasn't been taught yet.

8. ***Jeden Tag <u>meine Schwester geht</u> im Wald joggen.***
   a. I don't know why this part of the sentence is underlined, i.e. wrong.
   b. I know that the underlined part is wrong but I can't explain why.
   c. The underlined part is wrong because

      _____

      _____

   d. This feature hasn't been taught yet.

9. ***Ich habe <u>nicht</u> einen Freund in Toronto.***
   a. I don't know why this part of the sentence is underlined, i.e. wrong.
   b. I know that the underlined part is wrong but I can't explain why.

c. The underlined part is wrong because

_____

_____

d. This feature hasn't been taught yet.

10. Meine Freundin schenkt <u>die Frau</u> zum Geburtstag einen Kuchen.

a. I don't know why this part of the sentence is underlined, i.e. wrong.
b. I know that the underlined part is wrong but I can't explain why.
c. The underlined part is wrong because

_____

_____

d. This feature hasn't been taught yet.

11. Meine Mutter hat noch nie einen Grizzly hier in Kanada <u>geseht.</u>

a. I don't know why this part of the sentence is underlined, i.e. wrong.
b. I know that the underlined part is wrong but I can't explain why.
c. The underlined part is wrong because

_____

_____

d. This feature hasn't been taught yet.

Appendix E: Coding Features Metalinguistic Knowledge Test

Identified core features (expressed in grammatical terminology) used for data coding as agreed upon in the expert group

Driving question: Will the changes that would result from the explanations lead to a correction of the sentence (i.e. according to domain)? → yes = 1, no = 0

The correction versions did not have to be provided (as was stated in the instrument instructions)

1. ***Studenten an der University of Alberta <u>können kaufen</u> auf dem Campus viele Bücher.***

   Position of (main) verb is incorrect (since combined with modal verb, which has to be in 2nd position); main verb has to be pushed to the end / final position

2. ***Im Sommer 2016 <u>haben</u> meine Freunde und ich jedes Wochenende in die Disko gegangen.***

   Auxiliary / verb is incorrect; correct form: sind
   Gehen = action verb that requires the use of “sein” instead of “haben”

3. ***Weil ich <u>habe</u> viele Kurse, kann ich meine Freunde nicht sehen.***

   Verb “habe” needs to be moved to the end of the weil-clause because dependent clauses require the verb to be in final position

4. ***Ich habe <u>ein Bruder</u> und er hilft mir oft mit den Mathematikhausaufgaben.***

   „ein Bruder“ is the direct object and therefore requires the use of the accusative; “ein Bruder” is nominative

5. Meine Lehrerin kommt <u>in die Klasse jeden Morgen um 9 Uhr.</u>

Order of adverbs (of time and place) is wrong; time (jeden Morgen um 9 Uhr) has to be mentioned before place (in die Klasse)

6. Viele Menschen <u>aufstehen</u> sehr früh, wenn sie arbeiten müssen.
"Aufstehen" is a separable prefix-verb, "auf" has to move to final position of the independent clause

7. Sabine, Peter und Michael <u>fährt</u> am Nachmittag oft ins Fitnessstudio.

Lack of subject-verb agreement; subject refers to three people; verb form is wrong: fahren

8. Jeden Tag <u>meine Schwester geht</u> im Wald joggen.

In an independent clause, verb (= geht) has to be in second position, subject in third (in this case)

9. Ich habe <u>nicht</u> einen Freund in Toronto.

Nicht negates verbs; a noun (einen Freund) has to be negated with a form of kein (here: keinen)

10. Meine Freundin schenkt <u>die Frau</u> zum Geburtstag einen Kuchen.

Indirect object (die Frau) takes the dative case; "die Frau" is accusative and has to be changed to "der Frau"

11. Meine Mutter hat noch nie einen Grizzly hier in Kanada <u>geseht.</u>

Form of participle is wrong; it is a stem-changing and irregular (strong) verb; participle requires the use of "en" at the end

Appendix F: Coding Explanations Transparency and Opacity

The table provides an overview of combinations of modifier and head as identified in the data and how they were labelled. Examples are provided as well.

Combination of Modifier and Head	Label / Category
opaque + opaque Example: dative object	opaque
transparent + transparent Example: plural form	transparent
transparent + pseudo-transparent Example: main sentence	mostly transparent
pseudo-transparent + transparent Example: sentence structure	mostly transparent
opaque + transparent Example: verbal bracket	semi-transparent
transparent + opaque Example: separable prefix	semi-opaque
opaque + pseudo transparent Example: adverbial phrase	mostly opaque
pseudo transparent + opaque Example: direct object	mostly opaque

Appendix G: Coding Explanations Multi-Word Expressions

Multi-word expressions

In addition to individual terms, grammatical terminology often contains compound expressions, for example “verbal bracket” or “simple past”. When determining the level, or degree, of opacity or transparency, each term needs to be assessed individually but since they are a conceptual unit, they were counted, i. e. coded as such, for example, *verbal bracket* and *past tense.*

While “bracket” and “past” are considered to be transparent terms because the connection between signifier and signified is direct, “verbal” and “tense” are coded to be opaque based on the reasoning provided above (i. e. knowledge of Latin or French etymology, in the case of “tense”).

In case one term was opaque and one other one was transparent, the category was either “semi-transparent” or “semi-opaque”, depending on the head of the compound.

When the head of the compound was transparent, the term was coded as semi-transparent: *verbal bracket.* When the head was opaque, it was coded as semi-opaque: *past tense*. With English being a head-final language and the head determining the core meaning of the compound, modified by the preceding term, this decision was considered as valid.

In addition to the head being the determining factor for whether transparency or opaqueness was in the focus of the label (i. e. semi-transparent vs. semi-opaque), the overall dominance of a certain attribute was a decisive factor as well. For example, when

we have a combination of “opaque + pseudo-transparent”, “transparency” is in the head of the expression (e.g. adverbial phrase), which should lead to “transparency” being the dominant factor in labelling the term. However, since “pseudo” expresses a decreased level of transparency, and thereby a tendency towards opaqueness, the combination of opaque and pseudo-transparent leads to “mostly opaque” For terms that consisted of more than two components (e. g. *separable prefix verb* or *two-part verb phrase*), the categories head and overall dominance applied as well:

separable	***prefix verb***	
transparent	opaque + opaque	= mostly opaque
two-part	***verb phrase***	
transparent	opaque + pseudo-transparent	= mostly transparent

Appendix H: Representative Examples for Grammatical Explanations in Textbooks

The following examples can be considered as representative of the Latin-based notion of grammar to illustrate the nature of the metalanguage used in the beginner textbooks for German as a foreign language.

Example 1: The dative case

> **The dative case**
>
> The dative case has three major functions in German: it is the case of the INDIRECT OBJECT, it follows certain verbs, and it follows certain prepositions.
>
> In English the INDIRECT OBJECT is indicated in two ways:
>
> - Through word order: *The boy gives **the father** the plate.*
> - With a preposition: *The boy gives the plate **to the father.***
>
> In German this function is expressed through case and word order. You can determine the indirect object by asking *for whom* or *in reference to whom* (or occasionally *what*) the action of the verb is taking place.
>
> Der Junge gibt **dem Vater** den Teller *The boy gives the father the plate.*
>
> (Wie gehts, p. 82)

As the very first sentence reveals, the meaning of the terms “dative” or “case” is not explained but their three “major functions” are presented, which leads to a plethora of more Latin-based terms such as indirect object, prepositions, etc.

When the indirect object is introduced, the learners are confronted with the questions they need to ask in order to identify the object (for whom / in reference to whom or what) while the reference norm, i. e. what does “indirectness” refer to, is not addressed at all.

In this explanation we also find a contrastive approach to English by demonstrating how an indirect object is “indicated” in English; again, no explanation is provided but an exemplification including components of signal grammar (i. e. by putting the English equivalent in bold letters). However, the exemplification is not clear or straightforward. By just putting “the father” in bold letters when mentioning word order, it is difficult to derive what is distinct about this word order in contrast to other options.

The indication of an indirect object with a preposition is also unclear: not only the preposition “to” but the entire prepositional phrase “to the father” is in bold letters, which makes it difficult to identify the preposition in question. An explanation for what a preposition actually is and what it does is missing (that applies to the previous chapters of this particular textbook as well, which means the term is assumed to be part of the pre-existing metalinguistic knowledge of the learner).

Example 2: The perfect tense

> **Talking about the past: the perfect tense**
>
> In conversation, German speakers generally use the perfect tense to describe past events. The simple past tense, which you will study in **Kapitel 9**, is used more often in writing.
>
> | | |
> |---|---|
> | Ich **habe** gestern Abend ein Glas Wein **getrunken**. | I drank a glass of wine last night. |

Nora **hat** gestern Basketball **gespielt**. Nora played basketball yesterday.

German forms the perfect tense with an auxiliary (**haben** or **sein**) and a past participle (**gewaschen**). Participles usually begin with the prefix **ge-**.

[…]

The auxiliary is in first position in yes/no questions and in second position in statements and w-word questions. The past participle is at the end of a clause.

Hat Heidi gestern einen Film **gesehen**? Did Heidi see a movie last night?

[…]

Wann **bist** du ins Bett **gegangen**? When did you go to bed?

Although most verbs form the present perfect with **haben**, many use **sein**. To use **sein**, a verb must fulfill two conditions.

1. It cannot take a direct object.
2. It must indicate change or location or condition.

[…]

(Kontakte, p. 161)

This section of the perfect tense in German starts with an explanation of what it does: “to describe past events”. It is then contrasted with another past tense, namely the simple past, which is not explained further but only announced to be of interest in chapter 9 of the book. This explanation is followed by German examples including their English translation. As was true in the example above, the relevant portions, i. e. the two components the perfect tense in German consists of, are provided in bold letters. By providing an English translation, a contrastive approach is automatically implied. In that context, this decision is rather unfortunate because the German sentences resemble the present perfect tense in English.

When the auxiliary verbs “sein” and “haben” are introduced. they are only labelled as such, without explaining what an auxiliary verb is and what it does. The same is true for the introduction of a participle – only an example is provided and an indicator for how to recognize a participle, namely by the prefix “ge-”, which is not further explained or translated.

Example 3: Modal verbs

> **Modals**
>
> Modals express obligation (**sollen**), ability (**können**), necessity (**müssen**), permission (**dürfen**), and desire or preference (**wollen** / **mögen**). Modals are always conjugated and appear with another verb, the dependent infinitive. This infinitive is placed at the end of the sentence.
>
> […]
>
> Ich **will** Schriftsteller **werden**.
> I want to become a writer.
>
> Ich **kann** gut **malen**.
> The artist can paint well.

This last example also serves to demonstrate that, again, no explanation is provided but examples are given instead.

Furthermore, despite of expressing very different concepts such as obligation, ability, and so on, modal verbs share the common denominator of adding certain “modes” or nuances to the main verb. This underlying core idea of modal verbs is completely missing in this explanation.

As was true for indirect objects, dependent infinitives are mentioned but neither explicitly explained (only referred to as “another verb”) nor presented against the reference norm they are operating with – what does dependency / independency refer to?