Stability of Grammaticality Judgments in German-English Code-Switching

by

Jane Grabowski

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Approved April 2011 by the
Graduate Supervisory Committee:

Jeffrey MacSwan, Co-Chair
Daniel Gilfillan, Co-Chair
Carla Ghanem

ARIZONA STATE UNIVERSITY

May 2011

ABSTRACT

Code-switching, a bilingual language phenomenon, which may be defined as the concurrent use of two or more languages by fluent speakers is frequently misunderstood and stigmatized. Given that the majority of the world's population is bilingual rather than monolingual, the study of code-switching provides a fundamental window into human cognition and the systematic structural outcomes of language contact. Intra-sentential code-switching is said to systematically occur, constrained by the lexicons of each respective language. In order to access information about the acceptability of certain switches, linguists often elicit grammaticality judgments from bilingual informants. In current linguistic research, grammaticality judgment tasks are often scrutinized on account of the lack of stability of responses to individual sentences. Although this claim is largely motivated by research on monolingual strings under a variety of variable conditions, the stability of code-switched grammaticality judgment data given by bilingual informants has yet to be systematically investigated. By comparing grammaticality judgment data from 3 groups of German-English bilinguals, Group A (N=50), Group B (N=34), and Group C (N=40), this thesis investigates the stability of grammaticality judgments in code-switching over time, as well as a potential difference in judgments between judgment data for spoken and written code-switching stimuli. Using a web-based survey, informants were asked to give ratings of each code-switched token. The results were computed and findings from a correlated groups t test attest to the stability of code-switched judgment data over time with a p value of .271 and to the validity

of the methodologies currently in place. Furthermore, results from the study also indicated that no statistically significant difference was found between spoken and written judgment data as computed with an independent groups t test resulting in a p value of .186, contributing a valuable fact to the body of data collection practices in research in bilingualism. Results from this study indicate that there are significant differences attributable to language dominance for specific token types, which were calculated using an ANOVA test. However, when using group composite scores of all tokens, the ANOVA measure returned a non-significant score of .234, suggesting that bilinguals with differing language dominances rank in a similar manner. The findings from this study hope to help clarify current practices in code-switching research.

ACKNOWLEDGMENTS

Although only my name appears on the cover of this thesis, many people have contributed to its production. I am truly grateful to my fantastic thesis committee, whose valuable insights and guidance have made the completion of this project possible.

Dr. Jeff MacSwan, who co-chaired my thesis committee, first sparked my interest in the formal study of linguistic code-switching. Without his guidance, criticism, and encouragement, a far inferior product would have resulted. I consider myself incredibly fortunate to have had the opportunity to work with Dr. MacSwan and to him I give my deepest gratitude.

My co-chair, Dr. Daniel Gilfillan, is quite possibly one of the best teachers I have ever had. He has taught me to think critically and to continuously question how I can improve the relevancy of my work. I have been thoroughly fortunate to have an advisor who gave me the freedom to explore my own interests. Thank you so much for your amazing guidance!

I am incredibly grateful to Dr. Carla Ghanem for her valuable insights and encouragement and for giving both with such grace. Dr. Elly van Gelderen is quite the inspiration and I hope to produce work of her caliber some day. A very heartfelt thanks goes to Dr. Bryan Smith and Dr. Sarah Brem, who patiently explained various aspects of statistical analysis to me.

Last but not least, I would like to extend my gratitude to my family, who has been code-switching incessantly since I was born.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 Rationale for Code-Switching Research

Code switching (hereafter CS), a bilingual language phenomenon, occurs

frequently in language communities around the world in a variety of forms,

ranging from oral aspects to varying written traditions, such as song, advertising,

literature, and more recently with the expanse of the internet, in blogs and web

pages. Generally, it may be defined as the "ability on the part of bilinguals to

alternate effortlessly between their two languages" (Bullock & Toribio, 2009, p.

1). A bilingual is traditionally described as an individual possessing two or more

languages. Given that an estimated 6,000 languages are currently spoken in the

world (e.g. Crystal, 1987; Baker & Prys Jones, 1998), distributed over slightly

less than 200 countries, language contact on account of politics, natural disaster,

religion, culture, economy, education, and technology is inevitable. Many

opportunities for language contact present themselves within one's own country,

community, neighborhood, and family on account of the languages to countries

ratio. Consequently, the majority of the world's speakers are bilingual rather than

monolingual, while "one in three of the world's population routinely uses two or

more languages for work, family life, and leisure" (Wei, 2000, p.5). If people who

make irregular use of a foreign language were to be considered bilinguals, in

reality only a fraction of the world's population could be considered monolingual.

Given that more people in the world are bilingual than monolingual, CS is

invaluable as a study of cognition.

"While CS is viewed as an index of bilingual proficiency among linguists, it is more commonly perceived by the general public as indicative of language degeneration" (Bullock &Toribio, 2009, p. 1). The incompatibility of opinions between linguists and laypeople may be attributed to the disparity in the definition of *grammar.* While the general public understands grammar from a *prescriptive* perspective which mandates how language ought to be used, some linguists are interested in more objective, *descriptive* approach to grammars, "which represent speakers' unconscious knowledge of their languages as manifested in their actual linguistic behavior" (Bullock &Toribio, 2009, p. 1). The following excerpt is taken from a wildly popular German magazine column called "Zwiebelfisch" by Sebastian Sick reflecting the general public's attitude towards language contact in Germany.

> Die Präposition "in" vor einer Jahreszahl ist ein lästiger Anglizismus, der vor allem im Wirtschaftsjargon allgegenwärtig ist. Die deutsche Sprache ist jahrhundertelang ohne diesen Zusatz ausgekommen und braucht ihn auch heute nicht. Der Zweite Weltkrieg war nicht "in 1945" vorbei, sondern 1945. Ich wurde nicht "in 1965" geboren, sondern 1965. Die Formulierung "Der Film wird voraussichtlich erst in 2006 in die Kinos kommen" zeugt nicht nur von schlechtem Stil, sie ist außerdem länger als die korrekte deutsche Fassung, für die man das "in" ganz einfach streicht.[1]
>
> (http://www.spiegel.de/kultur/zwiebelfisch/0,1518,311727,00.html)

---

[1] The preposition "in" placed before a year is an annoying Anglicism, which has become ubiquitous in marketing jargon. The German language has managed for

Linguists, on the other hand, consider CS to provide a window into the systematic structural outcomes of language contact. "Further, the act of CS can be studied as a reflection of social constructs and of the cognitive mechanisms that control language switching" (Bullock &Toribio, 2009, p. 1). A bilingual who engages in CS is "privy to extremely subtle grammatical requirements" (MacSwan, 1999, p. 250) and has an inherent knowledge of more than one language systems and the parameters under which they may interact. In order to gain access to the interaction between varying language systems, researchers employ various experimental tactics, among those are grammaticality judgment tasks which probe a code-switcher's intuition about the grammatical acceptability of certain switches.

In current research, the use of judgment tasks has been criticized in that judgments are not made in a consistent, stable way and discrete judgments are not capable of capturing fine grades of grammaticality (Bader & Häussler, 2010). Given that CS is considered to be a sociolinguistically motivated spoken discourse strategy, privy to phonological restrictions, and rarely found in written texts, the use of written questionnaires for the assessment of acceptability presents a relatively formal and unnatural context, and is thought to skew the sentence ratings (Bialystok & Ryan, 1985; Greenbaum, 1977b). Although this view is largely motivated by monolingual inquiries, CS grammaticality judgment data

---

centuries without this addition, and doesn't need it nowadays either. World War II did not end "in 1945", but rather "1945". I was not born "1965", not "in 1965". The phrase "The film is expected to be released in 2006" shows not only poor style, it is also longer than the correct German version, which can be achieved by simply deleting the "in". (self-translated)

elicited in written form has yet to be compared to data from spoken judgment tasks.

In order to address this gap in literature pertaining to methodological approaches to CS, this thesis will attend to several conceptual and methodological questions. The principal objective of this study is to determine whether German-English bilingual informants' judgments for written tokens are stable over time. If the judgments are stable, can this generalization be assumed for the whole group or are there statistically significant differences between the judgments of German- or English-dominant bilinguals and balanced bilinguals? In a similar manner, are there measurable differences in judgment data when the informants are presented with a written or a spoken token?

In hopes of answering these questions, this study presented 3 groups of German-English bilinguals with various grammaticality judgment tasks which aim to uncover possible differences in judgment patterns of complex German-English sentences. The aim of this study is to contribute to the source of facts available about methodological practices in code-switching research, inform syntactic –theoretical debates, and to contribute to a model of bilingual language processing.

1.2 Definition of Code-Switching

Like most aspects of language contact phenomena, the study of CS often faces terminological issues and confusion. It comes as no surprise that researchers fail to agree not only on their research programs, but also on the terminology

which they employ. As such, one may encounter terms such as *code-switching*, *code-mixing*, *borrowing*, and *code-alteration* when reviewing literature pertaining to the language contact phenomenon in which two or more languages are used within a single discursive situation. Although the "term *code* is a relatively neutral conceptualization of a linguistic variety – be it a language or a dialect" (Boztepe, 2003, p.4), the term in itself is rarely addressed in most definitions. MacSwan (1999) defines CS as a "speech style in which fluent bilinguals move in and out of two (or conceivably more) languages" (p. 37). Milroy and Muysken (1995) assume a similar position, in which CS is "the alternative use by bilinguals of two or more languages in the same conversation" (p.7). MacSwan, Milroy, and Muysken assert CS as an umbrella term for varying bilingual behaviors. The multitude of definitions alludes to the difficulty of definitely characterizing the broad range of contact phenomena CS comprises.

> First, its linguistic manifestation may extend from the insertion of single words to the alternation of languages for larger segments of discourse. Second, it is produced by bilinguals of differing degrees of proficiency who reside in various types of language contact settings, and as a consequence their CS patterns may not be uniform. Finally, it may be deployed for a number of reasons: filling linguistic gaps, expressing ethnic identity, and achieving particular discursive aims, among others. (Bullock & Toribio, 2009, p. 2)

CS is distinguished from borrowing and the like in that the languages which are mixed remain phonologically distinct, whereas "borrowing involves the full

phonological and morphological integration of a word from one language into the other" (MacSwan & McAlister, 2010, p. 1). For example, the English word *shopping* has been adapted into the German language as a loanword to fill a particular semantic niche, as this word denotes a particular type of shopping, for all things non-grocery related. It has been adjusted to fit phonological and morphological restrictions as *shoppen*, the *-en* suffix indicating the infinitive form for German verbs.

1.3 Levels of Code-switching

CS may occur at varying levels; inter-sentential, intra-sentential, or intra-word. Essentially, CS occurring at the inter-sentential level (1) "involves switches from one language to the other between sentences," (Myers-Scotton, 1993, p. 3) and coordinate and subordinate clauses, whereas intra-sentential CS (2) involves "switching within the same sequence or sentence fragment." (Myers-Scotton, 1993, p. 4) The points at which CS may occur in intra-sentential switching are being debated, but the general consensus is that it may be seen at points where the syntax of the two (or possibly more) languages aligns within a sentence.

(1)   *Sophie hat   Angst* because she is stuck in a tree.

Sophie has   fear   because she is stuck in a tree.

'Sophie is scared because she is stuck in a tree.'

(2)   I love *meine*   *Katzen.*

        I love   my-AKK-PL   cats-AKK-PL

        'I love my cats."

1.4 Various Proposed Restrictions in Code-switching

Inquiries into the study of CS have traditionally been approached from one of two perspectives, which have been studied either in term of grammatical or sociolinguistic aspects. Grammatical or structural approaches in CS aim to identify syntactic and morphological characteristics of CS constructions. Sociolinguistic approaches, on the other hand, attempt to determine patterns of occurrence, as well as the discursive functions of CS.

Gumperz (1982) chose to view strategies in CS from a sociolinguistic perspective, stating that "detailed observation of verbal strategies revealed that an individual's choice of speech style has symbolic value and interpretive consequences that cannot be explained simply by correlating the incidence of linguistic variants" (p. vii). Labov (1971) described CS as "the irregular mixture of two distinct systems" (p.45) implying that there were no detectable restrictions between the languages used in CS. Gumperz approaches constraints on CS from a discourse strategy competence perspective but does not discount that the languages' grammars hold a role of some significance. Scholars interested in grammatical aspects of CS search for principles which account for contrasts between grammaticality such as (3) and (4).

(3)　　My cat　　　*trinkt*　　　*viel*　　*Milch*.

　　　　My cat　　　drinks-3PS　　much　milk-AKK-S

　　　　'My cat drinks a lot of milk.'

(4)　　*She　trinkt　　　　viel　　Milch.

　　　　She　drinks-3PS　　much　milk-AKK-S

　　　　'She drinks a lot of milk.'

With respect to grammatical aspects of CS, there have been a multitude of approaches attempting to describe universal restrictions which hold true for any two language pairs.

Early work on grammatical restrictions in CS began to uncover that "CS behavior is itself rule-governed" (MacSwan & McAlister, 2010, p. 3). The construction-specific constraints of CS explored by Gumperz (1967, 1970), Timm (1975), and Wentz (1977) soon led to a constraint[2] based approach, which posited "the kind of theoretical constraints developed in the contemporaneous syntactic literature to impose vertical limits on transformations and phrase structure" (MacSwan & McAlister, 2010, p. 3). Although constraints were posited by Chomsky (1964, 1965) and Ross (1967) to initially account for the over expressiveness of the generative-transformational approaches to a single grammar, CS-specific constraints were soon offered up by researchers.

---

[2] In the theoretical sense, a constraint is a principle or rule of grammar that bars one or another construction.

The most popular approaches were Sankoff and Poplack's (1981)

Equivalence Constraint[3] and *Free Morpheme Constraint*[4], Joshi's (1985) *Closed-*

Class Items Constraint[5], Di Sciullo, Muysken, and Singh's (1986) *Government*

Constraint[6], Mahootian's (1993) *Complement Relation Constraint*[7], and Belazi,

Rubin, and Toribio's (1994) *Functional Head Constraint.* Of particular relevance

to CS is Belazi et. al's model. The *Functional Head Constraint* states that "the

language feature of the complement f-selected by a functional head, like all other

relevant features, must match the corresponding feature of the functional head"

and "a code switch may not occur between a functional head and its complement"

(Belazi, Rubin & Toribio, 1994, p.228). Although constraints may hold up for a

specific language pair, copious amount of data and studies, such as Cantone and

Müller (2008) on their work on German-Italian CS, Liceras et al. (2008) with

English-Spanish CS, and MacSwan (2005) with Spanish-Nahuatl CS, have long

argued against the functional head constraint. Despite their respective

---

[3] Codes will tend to be switched at points where the surface structures of the
languages map onto each other. (Poplack, 1981)
[4] A switch may occur at any point in the discourse at which it is possible to make
a surface constituent and still retain a free morpheme. (Poplack, 1981)

[5] Closed-class items (e.g. determiners, quantifiers, prepositions, possessives, Aux,
Tense, helping verbs) cannot be switched. (Joshi, 1985)

[6] There is an anti-government requirement on CS boundaries. Chomsky (1995)
has since then suggested that government relation be considered in terms of
feature checking.

[7] The language of a head determines the phrase structure position of its
complements in code switching just as in monolingual contexts. (Mahootian,
1993)

contributions to the field of linguistics and CS, the above mentioned constraints have all been dispelled through numerous counter examples.

A constraint based approach to CS is problematic on several accounts. Chomsky (1955) proposed the notion of Universal Grammar (UG), which asserted that all languages have a common structural basis, which is determined by a limited set of rules. As such, it seems counter intuitive to introduce CS-specific devices and mechanisms or to assume a third grammar for CS. Pfaff (1979) echoes this sentiment and writes that "it is unnecessary to posit the existence of a third grammar to account for the utterances in which languages are mixed" (p. 314). Santorini and Mahootian (1995) further argue against a third grammar and assert that "codeswitching sequences are not subject to structural constraints beyond the general principles of phrase structure that monolingual sequences" (p. 3). Sankoff and Poplack (1981) also prefer a theory-driven analysis, positing that "the rules used to construct its constituents may be drawn at times from one monolingual grammar and at times from another" (p.12).

1.5 A Constraint-Free Approach to Code-Switching

Modern work in linguistics has approached syntactic theory with the goal of developing a model that can successfully generate all grammatical sentences within a language and reject all ungrammatical ones (Chomsky, 1957, p. 11-13). Rather than assuming a "third grammar," MacSwan proposes a minimalist approach to intra-sentential CS. MacSwan's thesis rests on Chomsky's Minimalist Program, whose model describes two central components inherent to

10

language. The first is a computational system, which is invariant across languages. The second component is the lexicon, which restricts the parameters observed to be unique in each language. As a result, phrase structure is dependent on the lexicon and is built derivationally by three basic operations *Select*, *Merge*, and *Move*. The *Select* operation "picks items from the lexicon and introduces them into the lexical array, an assembled subset of the lexicon used to construct a derivation" (MacSwan, 1999, p. 67). *Merge* organizes lexical items within the array into hierarchical syntactic objects or substructure, while *Move*, which is applied to the syntactic objects formed by *Merge*, is triggered by formal, uninterpretable features.

In his approach to CS, MacSwan argues that "if all syntactic variation is associated with the lexicon, as in the Minimalist Program, then code-switching may be seen as the simple consequence of mixing items from multiple lexicons in the course of derivation" (van Gelderen & MacSwan, 2008, p. 767). As a result, he posits a relevant theory of CS (5).

> (5) Nothing constrains code switching apart from the requirements of the
> mixed grammars. (146)

1.6 Verb placement in German and English

For the purpose of this study, I would like to focus on intra-sentential CS, with particular emphasis on verb movement in code-switched *weil-* and *because-* clauses. Given that intra-sentential CS is thought to occur at points where the syntax of the two languages is thought to align, if "German-English bilinguals

11

want to code-switch subordinate clauses, they need to resolve the problem of English being SVO whereas German finite verbs depending on subordinating conjunctions are generally placed in clause-final position (SOV)" (Eppler, 2004, p. 128).

The choice to use subordination for this study was based heavily on the amount of research already done on *weil-* clauses in German (Günther, 1993, 1996; Pasch, 1997; Uhmann, 1998; Farrar, 1999; Scheutz, 1998) and *weil-* clauses in German-English CS (Eppler, 1999; 2004). The interesting differences displayed between English and German syntax with respect to word order and verb placement are summarized as follows:

English is considered to be a subject-verb-object (SVO) language, which refers to its inherent sentence structure, where the subject comes first, followed by the verb in second position, while the object and all other dependents are placed third, such as in string (6). The position of the verb, in English, whether part of a main (6) or subordinating clause (7), is always after the subject and preceding all other dependents.

(6)   Alan **ate** the fish.

(7)   Emma is mad, because Alan **ate** the fish.

German on the other hand, is considered to be an SOV, or a verb final language, in generative grammar, despite allowing for both SVO (8) and SOV (9) type word orders in main clauses. In German main clauses, the non-finite verb is in final position while the finite verb remains in second position (9) or first position (10).

(8)   Alan   **isst**   den   Fisch.

        Alan   eats-3PS   the-AKK-M   fish-AKK-M

        'Alan eats the fish.'

(9)   Alan   **hat**   den   Fisch   **gegessen**.

        Alan   has-AUX-3PS the-AKK-M   fish-AKK-M eaten-PP

        'Alan ate the fish.'

(10)   **Hat**   Alan   den   Fisch   **gegessen**?

        has-AUX-3PS Alan   the-AKK-M   fish-AKK-M eaten-PP

        'Did Alan eat the fish?'

The word order found in subordinating clauses is thought to be the underlying

order for all sentences in German. "In tensed subordinate clauses, both finite (11)

and non-finite (12) (i.e. infinitival and participial) forms are final" (Hawkins,

1996, p. 132).

(11)   Ich   weiß, dass Sophie gerne   Mäuse   **fängt**.

        I   know that Sophie likes   mice   catches-3PS

        'I know that Sophie likes to catch mice.'

(12)   Ada spielte mit der Schnurr, weil   sie ihr **gefallen hat**.

        Ada played with the string because   it   her pleasing has

        'Ada played with the string because she found it pretty.'

The verb is also final in non-tensed German subordinate clauses. (13)

(13)   Emma freut sich darauf,   morgens ihr Futter **zu essen**.

        Emma excites herself for that tomorrow her food to eat

        'Emma looks forward to eating her food in the morning.'

Additionally, separable elements on complex verb forms, such as *mitspielen* (14) and *aufessen* (15), assume the same word order as the non-finite verb forms in (9)-(13).

(14)   Ada und Emma **spielen** heute   nicht **mit**.

Ada and Emma play   today   not   with

'Ada and Emma aren't playing along today.'

(15)   Sophie hat das Futter **aufgegessen**.

Sophie has the food ate-up

'Sophie ate up the food.'

Given the SOV nature of German syntax in subordinating clauses, contrasted with the preference for SVO structures for dependent clauses in English, exploring the boundaries of CS between the conjunctions *because* and *weil* and dependent strings in a contrasting language is of extraordinary interest to the inquiry of CS and may inform how constraints of this nature are resolved. In addition, eliciting grammaticality judgments for sentences of this nature will further inform the body of knowledge already made available by Eppler (2004).

While Chapter 1 gave a rationale for research in CS, Chapter 2 will give an outline of the goals of linguistic theory and how they may be achieved by collecting data through the elicitation of grammaticality judgments. Chapter 3 gives an outline of various methodological aspects applied in this particular thesis, while Chapter 4 gives descriptive and inferential statistics of the results. The study is concluded in Chapter 5 with suggestions for further research.

Chapter 2

LITERATURE REVIEW

2.1 Goals of Linguistic Theory

The goals of generative linguistics, as a subfield of cognitive psychology

and neuroscience, aim to answer a series of questions relating to what constitutes

knowledge of a language. In *Aspects of the Theory of Syntax* (1965), Chomsky

clarifies what he considers "knowledge of a language" and likens it to linguistic

competence.

> We thus make a fundamental distinction between *competence* (the
>
> speaker-hearer's knowledge of his language) and *performance* (the actual
>
> use of language in concrete situations). Only under the idealization set
>
> forth in the preceding paragraph is performance a direct reflection of
>
> competence. In actual fact, it obviously could not directly reflect
>
> competence. A record of natural speech will show numerous false starts,
>
> deviations from rules, changes of plan mid-course, and so on. The problem
>
> for the linguist, as well as for the child learning the language, is to
>
> determine from the data of performance the underlying system of   rules
>
> that has been mastered by the speaker-hearer and that he puts to use in
>
> actual performance. Hence, in the technical sense, linguistic theory is
>
> mentalistic, since it is concerned with discovering a mental reality
>
> underlying actual behavior. (p.5)

The knowledge of a particular language system is considered to be a (mental)

grammar. Every speaker of a language has acquired a tacit set of rules, in contrast

to the prescriptive grammar rules one may have learned in school, which allow him or her to utilize that language, whether through comprehension or production. In addition to seeking to identify the knowledge of a language within a speaker, generative linguistics attempts to identify the range of all human languages which we are capable of encoding, learning, and using. A normally developing child's ability to learn any language in a given linguistic environment regardless of place of origin, coupled with time and input constraints, suggests that there are properties, termed Universal Grammar (UG), which are common to all human languages. This hypothetical notion of UG, which allows for variation between languages, also asserts characteristics which are common to all languages. This could be likened to the genetic code, which ensures that a normally developing human has a set number of limbs, which are very similar in appearance, but not identical from one organism to another. Although UG is controversial, "many generative linguists concur with Chomsky's view that our mental grammar distinguishes at least two kinds of strings: those that are possible sentences of our language and those that are not." (Schütze, 2010, p. 2) All speakers of German would agree that sentence (16) is a well-formed, possible sentence of German, while string (17) is not.

(16)   Die Katze   ist   im   Baum.

  the cat-3PS-NOM   is-3PS   in the tree

  'The cat is in the tree.'

(17)   *Baum im   ist   Katze die.

        tree   in the is   cat   the

        'Tree in the is cat the.'

Furthermore, the distinction between 'well-formed' and 'ill-formed' strings is quite notably different from the distinction made between strings which are interpretable and those which are not privy to interpretation. Although a speaker of German could potentially guess the meaning of the ill-formed string (17), it is considered to be uninterpretable. In contrast to string (17), string (18), albeit also considered to be ill-formed, is closer to a well-formed sentence of German, remaining interpretable, similar and close in meaning to string (16).

(18)   *Die   Katze   bist   im   Baum.

        the   cat-3PS-NOM   is-2PS   in the tree

        'The cat are in the tree.'

Linguistic theory, which has devoted its attention to differentiating between grammatical strings, such as (16), and ungrammatical strings, such as (17) and (18), has paid "virtually no attention to the question of which strings can be assigned a consistent interpretation or how the aforementioned notion of closeness might be reified" (Schütze, 2010, p.2). Although all speakers of German would agree on the grammatical statuses of strings (16)-(18), the grammaticality of certain strings is ambiguous, in that speakers of the same 'language' will disagree whether a string is possible. The dissonance in judgment between speakers of a language may be systematic, often aligning on geographic or socio-cultural axes. Consider for example, that in standard German "causal clauses introduced by the

17

conjunction *weil* display subordinate, verb-final word order" (Miller, 2010, p.4),

such as statement (19).

> (19)   Die Katze sitzt im   Baum, weil   sie Vögel **fangen will**.
>
>   the cat   sits in the tree   because   it   birds catch wants
>
>   'The cat sits in the tree because it wants to catch birds.'

In southern Germany and Austria, however, the phenomenon of *weil* followed by

main clause (verb-second) word order has made its presence known in spoken

German (Küper 1991; Watzinger-Tharp 2006; Scheutz 2001), as illustrated in

statement (20).

> (20)   Die Katze sitzt im Baum,   weil   sie **will** Vögel **fangen**.
>
>   the cat   sits in the tree   because   it   wants birds catch
>
>   'The cat sits in the tree because it wants to catch birds.'

Although the study of CS also contributes in determining a person's knowledge of

language and mental grammar, "studies of CS are mainly concerned with the

nature of the constraints governing CS and its processing" (Gullberg, Indefrey &

Muysken, 2009, p. 21). Given that the focus is typically on the structural and

semantic relationship between phrases and on the linguistic constraints governing

those switches, "the methods used to study CS…are almost exclusively in the

domain of free production" (Gullberg, Indefrey & Muysken, 2009, p. 21). As a

consequence, that the study of CS is approached as a production phenomenon,

and the overarching methodological problem centers around how to induce,

manipulate, and replicate natural CS without compromise. As such, the role of

reliably collecting naturalistic data plays an important role in CS research.

2.2 Role of Naturalistic Data for the Study of Code-Switching

Linguists are privy to both naturalistic and experimental data. Naturalistic data consists of recorded and transcribed speech and is considered a source of positive evidence. "Works addressing the grammar of code-switching in bilingual speech have made use of a wide variety of methodologies, chief among these, interviews and naturalistic recording" (Toribio, 2001a, p. 405). Corpus data, or a collection of often naturalistic language data, is beneficial for the study of CS in a number of ways. Typically, naturalistic language data is left unbiased by the researcher. Corpora may also provide tokens of language production which were uttered while the speaker was not consciously focused on their formation. Naturalistic data is especially useful in that it allows for the exploration of different phenomena which are not testable as individual sentences need to be contextualized in order to be observed. While observations of natural speech are a fantastic source of data because they reflect actual language use, naturalistic data is met with a series of limitations:

1. Cost. Recording and transcribing a corpus of speech is costly, complicated, and time-consuming. Additionally, a corpus of bilingual speech, if it is to be used in CS research, needs to be phonologically annotated since "phonological systems cannot be mixed, and this property has the effect of creating discrete lexicons for the languages known to a bilingual person" (van Gelderen & MacSwan, 2008, p. 768).

2. Availability. Although there are some publically available corpora of bilingual speech, such as the BilingBank Corpora

(http://talkbank.org/data/BilingBank/), very few corpora on which CS studies are based are accessible.

3. No negative evidence. Although positive evidence is invaluable in support of certain linguistic theories, potentially providing confirmation of grammaticality, only negative evidence can confirm whether a certain string is ungrammatical. Both positive and negative evidence are essential to empirically verifying the validity of a proposed theory.

4. Problem of induction. "Codeswitching researchers committed to naturalistic data may assume that the absence of a specific pattern implies that it is grammatically impossible; however, we cannot confidently assume that the absence of a form in naturalistic data means that the structure is not permitted" (MacSwan & McAlister, 2010, p. 7). Perhaps a certain structure is barred from occurring, and as a result, it is absent from the corpus. However, it is also quite likely that the corpus was not extensive enough to account for that particular structure. "Thus, one cannot infer from the absence of a structure that such structures cannot occur, and waiting for relevant data to confirm or disconfirm theories narrowly focused on specific structures may lead to a perpetual disappointment. Elicitation of data provides a convenient and efficient shortcut: we can simply ask bilinguals who codeswitch whether a particular structure they might say in some context or another" (MacSwan & McAlister, 2010, pp. 7-8).

5. Problem of unidentified performance errors. Speech errors further complicate the role of naturalistic data as evidence for or against linguistic theory. Since "researchers have estimated the range of error rate in typically developing mature speakers to be as high as 10%" (Brown, 1973; Goodluck, 1991; Reilly, Marchman & Bates, 1998), one would expect a bilingual speaker to be privy to the same error rate as a monolingual when speaking. Naturally, when recording actual spoken language, it will be riddled with performance "errors of the sort we all make every day, errors we often recognize ourselves as inconsistent with our knowledge of language immediately after producing them" (MacSwan & McAlister, 2010, 10). Performance mistakes frequently occur when a speaker is forming a sentence while changing the thought process. As a result, strings such as (21) will be elicited in English, despite being permissible by the grammar.

     (21)   How much apples are left?

6. Fragmented speech. Spoken discourse is more dynamic and immediate, and as a result, fragmented.

On account of the high cost, lack of availability, absence of negative evidence, problem of induction, unidentified performance errors, and the fragmented nature of speech, observational research techniques fail to answer all the questions, and "more specific controlled experiments are called for to determine the contributions of the various factors involved" (Gullberg, Indefrey & Muysken, 2009, p. 26). Naturalistic data may also "be of limited value in the study of

linguistic competence, as it yields data that reflects the speakers' competence only indirectly" (Toribio, 2001a, p. 406). Despite some researchers' efforts to study solely the grammatical features of CS through naturalistic evidence, like much research in bilingualism, it is heavily reliant on social linguistic variables. Self-reports about bilingual speech are frequently unreliable. Bilinguals have difficulty remembering which language they used in a given speech situation, and many claim that they do not code-switch, only to be confronted with recordings of their speech riddled with CS occurrences. In contrast, speakers may consciously refrain from CS when being observed or recorded on account of social stigmas. MacSwan (1999) encountered various problems when attempting to collect data for his dissertation on Spanish-Nahuatl CS.

> "I quickly learned from working with Jesus that he had very negative attitudes toward code switching. When I asked him to express judgments on particular code-switched sentences he reacted with great discomfort. He believed that his language was losing ground among its people and the mixing of Spanish and Nahuatl was a great political disservice to the Aztec community." (p. 99)

As a result, the participant was excluded from the study, resulting in an extremely limited sample size.

2.3 Grammaticality Judgment Tasks

Prior to the advent of generative linguistics, researchers had rejected the utility of and access to a speaker's internal states. "Traditionally, concrete speech events, i.e., naturally occurring written or spoken utterances, were taken without further ado as the only relevant source of linguistic data" (Kepser & Reis, 2005, p. 2). Since the late 1950s however, linguistic intuitions have been an important source of evidence in the constructions of grammars, upon which many linguistic theories rely. In 1996, Schütze published his *Empirical Base of Linguistics*, which demonstrated "that the absence of a methodology of grammaticality judgments in linguistics constitutes a serious obstacle to meaningful research" (xi) and argued for a systematic approach to eliciting and collecting judgments. Since then, researchers have increasingly addressed the necessity of controlling variables in the collection of speaker judgment data in order to obtain more reliable data. In response to this, Schütze (1996) gives 4 key reasons for the use of grammaticality judgments in linguistic research.

1. "By eliciting judgments, we can examine reactions to sentence types that might occur only very rarely in spontaneous speech or recorded corpora" (p. 2). This is a particular problem for CS research, where corpora of a particular language pair are almost impossible to find. In her article *Word Order in German-English Mixed Discourse* (1999), Eppler discusses the asymmetrical distribution between the conjunctions of reason *weil* and *because* in the 8.5 hour conversational

23

transcribed corpus she used, despite an almost even distribution between German and English use.

2. "A second reason for using grammaticality judgments is to obtain a form of information that scarcely exists within normal language use at all-namely negative information, in the forms of strings that are not part of the language" (p. 2). Eppler's corpus data on CS of *weil* and *because* as summarized in Table 1 would lead us to believe that switches with *weil* between a English parent-English dependent, English parent-German dependent, and German parent- English dependent are ungrammatical on account of their absence in the corpus.

	parE-depE	parE-depG	parG-depG	parG-depE	total
because	86	5	16	6	123[8]
weil	0	0	59	0	59

Table 1. Language of parent and dependent of because and weil

If naturalistic data is solely considered, string (22), (23), and (24) are ill-formed.

(22)   I'm broke, weil   my cat **drinks** a lot of milk.

I'm broke because   my cat drinks a lot of milk

'I'm broke because my cat drinks a lot of milk.'

---

[8] This table was taken from Eppler (2004). In reality, the total column for *because* should add up to 113.

(23)　　　I'm broke, weil　　meine Katze　viel　Milch **trinkt**.

　　　　　I'm broke because　my　cat　a lot　milk　drinks

　　　　　'I'm broke because my cat drinks a lot of milk.'

(24)　　　Ich esse kein Fleisch, weil　I **am** a vegetarian.

　　　　　I　eat　no　meat　because I am a vegetarian

　　　　　'I don't eat meat because I am a vegetarian.'

3. "When one is merely observing speech, it is difficult to distinguish reliably slips, unfinished utterances, and so forth, from grammatical production" (p. 2). While consulting the Texas German corpus (http://www.tgdp.org/) it became increasingly evident that speakers would start multiple sentences before finishing one, obscuring the speaker's underlying competence. Additionally, speech production is full of errors given memory limitations. This imposed "cognitive burden" may be relieved by the utilizing grammaticality judgments.

4. Grammaticality judgments "minimize the extent to which the communicative and representational functions of language skill obscure our insight into its mental nature" (p. 2). Albeit somewhat controversial, this fourth claim assumes that constructing arbitrary sentences for judgment purposes structural, albeit functionless properties of language.

Critics of judgment data, however, assert that "not only is the elicitation situation artificial, raising the standard issue of ecological validity, but the subject is being asked for a sort of behavior that, at least on the face of it, is entirely

different from everyday conversation" (Schütze, 1996, p. 4). This is problematic

on several accounts. Bever (1970) writes that "linguistic intuitions do not

necessarily directly reflect the structure of language, yet such intuitions are the

basic data the linguist uses to verify his grammar" (p. 346). Grimshaw and Rosen

(1990) echo Bever's sentiment and state that "performance on the standard

linguistic task of making grammaticality judgments cannot be equated with

grammatical knowledge" (p. 188), especially given our inexplicit knowledge of

language and its grammar. Myers-Scotton (2006) contends linguists need to

devise experiments that mimic natural CS. Bentahila and Davies (1983) argue that

negative grammaticality judgments in CS are not indicative of syntactic

constraints. Consequently, Bever (1970), Birdsong (1989), and Gleitman &

Gleitman (1979) suggest that linguists are constructing grammars based on

linguistic intuition, rather than grammars relying on production and

comprehension. However, as speakers of a language we are continuously judging

the speech of others, so that we can adjust our register. In addition, bilinguals not

only adjust their register, but also the language they choose to use in a given

language situation.

Despite voicing his concerns about grammaticality judgments, Bever

(1970) adds that "rather than rejecting linguistic study, we should pursue the

course typical of most psychological sciences; give up the belief in an "absolute"

intuition about sentences and study the laws of the intuitional process itself" (p.

346). Grimshaw and Rosen also soften their criticism and add that it is not

impossible to deduce grammatical knowledge from grammatical judgments. "The

inevitable screening effects of processing demands and other performance factors do not prevent us from establishing the character of linguistic knowledge; they just make it more challenging" (p. 217) and "require inferential reasoning, sometime of the highly abstract sort" (p. 188). "Consistency among speakers of similar backgrounds, and consistency for a particular speaker on different occasions is relevant information. The possibility of constructing a systematic and general theory to account for these observations is also a factor to be considered in evaluating the probable correctness of particular observations" (Chomsky, 1964, pp. 79–80). "While grammaticality judgments offer a different access path from language use to competence, they are themselves just another sort of performance (Birdsong, 1989; Levelt et al, 1977; Bever, 1970b, 1974; Bever & Langendoen, 1971), and as such are subject to at least as many confounding factors as production, and likely even more.

The reliability and stability of judgment data is frequently criticized on account of instability in contemporary literature. Even Chomsky, who suggested linguists use judgments, finds that introspective judgments are not "sacrosanct and beyond any conceivable doubt. On the contrary, their correctness can be challenged in various ways, some quite indirect" (p. 79). Among countless numbers of concerns, Schütze (1999) finds that "linguists are not trained in methods for getting reliable data and determining which of two conflicting data reports is more reliable" and often refrain from imposing "any of the standard experimental control techniques, such as random sampling of subjects and stimulus materials or counterbalancing for order effects" (p. 4). There seems to be

a tolerance for informal methods in judgment collection, in which quite often, the sole subject is the linguist himself.

Doubts about the stability of judgment data are no stranger to linguistic literature however. Labov (1972b, 1975), Ross (1979) and Stokes (1974) found significant variability and disagreement in grammaticality between informants. Caroll, Bever, & Pollack (1981), Nagata (1988), and Snow & Meijer (1977) have found that individuals will give a different judgment for the same token on different occasions or under different conditions. Additionally, sometime the judgments given by informants do not agree with the actual linguistic behavior of the individual. Despite the skepticism regarding the stability of judgments and consequently the validity of judgment data, "judgments of "grammaticality" have found increasing use in the psychological literature" (Cowart, 1997, p. 5).

Although there is a growing awareness among researchers for the need for systematic data collection, many researchers working with CS are lagging behind.

Chapter 3

RESEARCH METHODS

It is quite evident that judgment data from grammaticality judgment tasks is important for theoretical argumentation in CS theory. Within the last decade numerous amount of journal articles, dissertations, and books on CS have been published heavily relying on grammaticality judgment data. "Traditionally, grammaticality and acceptability tasks are written off-line tests probing participants' grammatical knowledge. In bilingual studies, participants must respond by indicating whether a sentence with a particular type of switch is grammatical or not, or indicate its degree of acceptability on Likert scales" (Gullberg, Indefrey, & Muysken, 2009, p. 31). Introspective judgments of grammaticality, albeit providing voluminous evidence for syntactic theory, have been heavily scrutinized in recent literature. "Two prominent points of criticism are, first, that such judgments are not made in a consistent way, and, second, that discrete judgments are not capable of capturing fine grades of grammaticality" (Bader & Häussler, 2010, p.273). Despite the popularity of this experimental task, there are several identifiable problems with the collection of judgment data in research on CS, some of which the following research questions will hope to address.

3.1 Research Questions

This study investigates the stability of grammaticality judgments of individual code-switched strings. While the study is comprised of three groups of

German-English bilinguals, it hopes to be universally applicable to data collection in bilingualism and CS.

The primary research questions of the study are

1. Are German-English bilinguals' responses to individual code-switched written sentences stable over time?

2. How does the stability of German-English bilinguals' responses to individual code-switched sentences compare when the tokens are presented visually in contrast to being presented aurally?

3. Do German L1, English L2 bilinguals give different grammaticality judgments than English L1, German L2 or simultaneous bilinguals?

3.2 Selecting Participants

The careful construction of a study may begin with a variety of ways. The selection of informants for grammaticality judgment tasks in CS is especially important. Although there is gradience in bilingual proficiencies, and bilinguals of various proficiencies are known to code-switch, MacSwan (1999) asserts that

> "consultants used in any study of intrasentential code switching should be native bilinguals, relatively evenly dominant in both languages, have actively used both languages since infancy, have had continues, sustained exposure to both languages, and appear to have generally high verbal fluency in both languages." (p. 98)

In like manner, several studies on child and adult bilingualism support that competence in both languages is imperative for systematic, rule-governed CS

(McClure, 1981; Rakowsky, 1989; Meisel, 1994; Bhatia & Ritchie, 1996). Albeit a fantastic idealization, determining whether a participant is a balanced bilingual is difficult and messy, "because the bilingual's skill may not be the same for both languages at all linguistic levels" (Romaine, 1995, p. 12). Regardless, "monolingual-like control of two languages over all aspects of linguistic knowledge and use within all domains is rare, if possible at all" (Bullock & Toribio, 2009, p. 7).   In addition, Valdés (1981) emphasizes that CS data, whether naturalistic or judgment-task in nature, can only be collected from a bilingual community which values CS. Naturally, positing such stringent restrictions in the selection of informants results in a lack of statistically significant possible participants. As a result, MacSwan (1999) used 3 informants for his study of Spanish-Nahuatl CS.

For the purpose of this study, 3 groups of informants were selected. All 3 groups consist of graduate students in various Germanic Studies departments at research universities across the United States. This population of informants was chosen in order to satisfy most of the selection criteria listed in MacSwan (1999) and is based on several assumptions about the informants' proficiencies in both German and English, given that no proficiency measure was applied. Although arguably of varying levels of proficiency in both German and English, all speakers are assumed to be of at least near-native proficiency, as well as highly verbally fluent, given their esteemed positions. In addition, the informants continue to have sustained exposure to both languages, and are part of a bilingual community. The chosen participants, however, do not fulfill the selection criteria

31

in that only a small number have grown up in a bilingual household and have been actively using both languages since childhood. In a study exploring the effect of age of onset of exposure in bilingual judgment data for Slovak-English CS, McAlister (2010) shows "that early and late sequential bilinguals tend to share the same intuitions about Slovak-English CS as simultaneous bilinguals" (p. 240). In order to overcome the drawbacks of not finding the perfect balanced bilingual participant, the importance of a relevant sample size ought to be brought to light.

3.3 Sample Size

The importance of determining a sample-size in planning a statistical study, albeit difficult, cannot be overstated. "Sample size is important for economic reasons: An undersized study can be a waste of resources for not having the capability to produce useful results, while an over-sized one uses more resources than are necessary" (Lenth, 2001, p. 188). For studies in which an experiment may be repeated *ad infinitum*, sample size determination often relies on studying the power of a test of hypothesis. Studies in CS, unfortunately, are often met with a limited number of qualified participants. Cowart (1997) asserts that

> "for acceptability phenomena of the magnitudes relevant to many
> contemporary issues in syntactic theory, the minimum reasonable
> experiment will use eight or more informants. Some phenomena can be
> detected by smaller samples, and some phenomena will require larger

informant groups, but few phenomena of current interest can be adequately described with fewer than eight informants." (pp. 83-84)

Although judgments of grammaticality have been shown to be highly stable in certain monolingual tasks when collected via appropriate methods (Chomsky & Lasnik, 1977), many judgment phenomena are not stable enough so that a single informant can be reliably representative of the entire speech community. Eight informants may form a large enough sample size to give consistent judgments on obvious syntactic violations, such as string (25).

(25)   Ich habe kein Geld,   weil   my cat a lot of milk **drinks**.

        I have no money   because   my cat a lot of milk drinks

        'I don't have any money because my cat drinks a lot of milk.'

CS string (26), on the other hand, would require a more robust number of informants, since (V2) or main clause order has been increasingly occurring in German causal clauses, but is not considered acceptable by all speakers of German.

(26)   Ich habe kein Geld, because meine Katze **trinkt** viel Milch.

        I have no money because my   cat   drinks a lot milk

        'I don't have any money because my cat drinks a lot of milk.'

Most suitably "the number of informants required is determined by the stability of the phenomenon itself" (Cowart, 1997, p. 82).

For the purpose of this study, each survey was sent out to 300 potential participants. Out of the 300, 50 informants responded to survey A which consisted of written tokens, 34 responded to survey B which was identical in nature to

survey A, and 40 responded to Survey C, which was identical to surveys A and B but consisted of spoken and recorded tokens the participants could hear, as they were embedded as sound files in the survey site. Based on this, 3 major groups may be identified. Group A consisted of 50 participants given the written survey. An identical, although randomized survey was sent out to all of the participants of Group A 2 weeks after the initial test for a retest. The informants that responded a second time, which were 34 in number, were designated as Group B2, and their initial responses were subset from Group A for comparison. Group C was an entirely different group of participants and they were given a spoken version of the written grammaticality judgment tasks.

3.4 Experimental Setting

When collecting judgment data, the experimental setting ought to be considered. Each informant comes to the judgment task as an individual, with a unique attitude and opinion about language and CS, a range of competencies in both languages, and differing intellectual competency. Despite a need for standardization in CS research, it paradoxically appears that experiments need to be customized in order to accommodate the variability in informants. "Every experiment has to negotiate a balance between the standardization of materials and procedure required to meet scientific goals and the individual customization needed to elicit the relevant sort of uniformity in human performance" (Cowart, 1997, p. 86). Judgment tasks may be elicited in the form of a survey which may be distributed online without the presence of the researcher, or in a lab or field

setting, where the investigators are present. Although distributing a survey online could potentially fetch more participants, it makes it difficult to customize the experiment to achieve what Cowart refers to as "uniformity in human performance" (p. 86). An increase in sample size, however, could mask or normalize the variation which could be expected on account of the variability in the linguistic competence or attitude of the informants. The participants of Group A (N=50), B (N=34), and C (N=40) for this particular study were contacted via an E-Mail, provided in Appendix A and Appendix B respectively. The choice to distribute an online survey rested solely in a lack of local German-English bilinguals who could serve as potential informants.

Conducting an experiment in a laboratory or field setting is beneficial in that the investigator may "tailor the process to the needs of the individual informant" (Cowart, 1997, p. 88). The informant has the opportunity to ask for clarification during the course of the session and may give the researcher feedback and perspective which can be used as guidance for improving the procedure. In addition, a more clear set of instruction about the tasks, as well as a training period with training tasks may be implemented for the informants to reduce unnecessary noise and error variance. This is particularly important when asking participants to elicit judgments on code-switched statements, since CS is thought to be largely sociolinguistically motivated as a discourse strategy (Auer, 1998; Labov, 1972a; Gumperz, 1976, 1982; Gardner-Chloros, 2009). This sentiment is echoed by a participant from Group A, who gives the following statement about CS.

"For most of the code-switching I both do or hear, there is some sort of

reason for the switch, either because of cultural associations, what I

learned first, what other people say a lot, whether I am quoting or referring

to a statement that someone else has made, etc."

In order to alleviate anxieties about giving a judgment, the presence of the

investigator to explain what is desired of the participant, and to conduct a short

training session, proves to be invaluable.

Since the experimental setting did not allow for an investigator to explain

what was desired from the informant, a brief priming subsection was included to

ensure that the informant understood what was being asked. The priming section,

which may be found in Appendix D, included 3 German monolingual tokens, 3

English monolingual tokens, and finally 3 German-English CS judgment tasks, all

of varying levels of potential acceptability. Out of each group of tokens, one

statement was intended to be fully grammatical, one ungrammatical, and one

statement was included to be of questionable grammaticality. The training session

verified that the subjects understood the scale and were comfortable completing

the online survey. In addition, the warm-up materials eliminated potential error

variance and unnecessary noise.

3.5 Instructions

As a result of evidence from grammaticality tasks becoming increasingly

necessary and relevant to linguistic theory, the quality of various types of

linguistic evidence ought to be scrutinized and evaluated. When we are asking

36

participants to give a judgment, we need to establish a certain level of confidence that the participant is actually doing what they have been instructed to do. The role of the instructions in eliciting quality data is imperative, given that "naïve speakers, left without proper guidance, may stray far from what we intend to be asking them" (Schütze, 2005, p. 457). In a study by Maclay and Sleator (1960), native English speakers in beginning rhetoric classes at the University of Illinois were asked "Do these words form a grammatical English sentence?" Out of the 21 students asked, three affirmed that string (27) was a grammatical sentence of English, while only four participants judged string (28) to be grammatical.

(27)   Label break to calmed about and.

(28)   Not if I have anything to do with it.

It seems evident that subjects should not be expected to know what is meant by *grammatical* or *acceptable*, since "it is not even clear that linguists agree among themselves as to what exactly is supposed to count towards grammaticality" (Schütze, 1996, p. 132). In an experiment examining the role of instructions in judgments regarding *that*-trace effects, Cowart (1997) split his subjects into two groups. Although the judgment experiment remained the same, one group received instructions asking for a "gut reaction, not on rules you may have learned about what is "proper" or "correct in English," while the other group was asked whether or not the sentence would be accepted by a professor teaching a 400-level English course. Cowart (1997) found that "difference in instructions did not produce any difference of pattern that seemed to matter to linguistic theory" (p. 57). This may be on account of several factors, among them the stability of the

that-trace effect or that regardless of instruction, subjects are only capable of one judgment type.

The role of instruction in judgment experiments pertaining specifically to CS, however, remains unexplored and is strongly recommended for further research. For the purpose of this study, I utilized the following instructions which Dussias (personal communication, February 19, 2011) implements when she runs grammaticality judgments.

"You will read (hear) sentences silently one by one on the computer screen. After each sentences, you will be asked to indicate whether the sentence you just read (heard) sounds OK to you. By "sounds OK," this is what we mean:

(1) if it is something that you would hear (or have heard) other code-switchers in your community say

(2) if it is something that you think you yourself say

(3) if it is something that you think you would say"

Dussias avoids using the terms *grammaticality* or *acceptability* making the instructions accessible to most subjects. She added that she did not give the informants examples of what might be acceptable or unacceptable "as not to bias or color their judgments."

3.6 Demographic Data

Although "for many syntactic investigations, there will be no need to collect demographic data" (Cowart, 1997, p. 90), a demographic data survey

38

modeled after a similar survey used by Dussias (2003), provided in Appendix C, was included on account that variation in the language demographic is not only relevant to this particular study, but most studies done in CS. In a study by Aguirre (1977), balanced Spanish-English bilinguals exhibited greater sensitivity to patterns in CS than English-dominant or Spanish-dominant bilinguals. Another study by Belazi (1991) suggests that fluent Arabic-French bilinguals are more sensitive to CS constraints than less balanced bilinguals. As a result, the degree of bilingualism and other as of yet unexplored demographic data must be taken into consideration for various investigations into grammatical aspects of CS.

Group A

In which language do you have greater SPEAKING proficiency?		
Answer Options	Response Percent	Response Count
German	22.0%	11
English	68.0%	34
Same for both English and German	10.0%	5
		50

Figure 1. Speaking proficiency language dominance for Group A

Group A

In which language do you have greater READING proficiency?		
Answer Options	Response Percent	Response Count
German	18.0%	9
English	66.0%	33
Same for both German and English	16.0%	8
		50

Figure 2. Reading proficiency language dominance for Group A

Group B

In which language do you have greater SPEAKING proficiency?		
Answer Options	**Response Percent**	**Response Count**
German	21.0%	7
English	68.0%	23
Same for both English and German	11.0%	4
		34

Figure 3. Speaking proficiency language dominance for Group B

Group B

In which language do you have greater READING proficiency?		
Answer Options	**Response Percent**	**Response Count**
German	17.5%	6
English	65.0%	22
Same for both German and English	17.5%	6
		34

Figure 4. Reading proficiency language dominance for Group B

Group C

In which language do you have greater SPEAKING proficiency?		
Answer Options	**Response Percent**	**Response Count**
German	17.5%	7
English	72.5%	29
Same for both English and German	10.0%	4
		40

Figure 5. Speaking proficiency language dominance for Group C

Group C

In which language do you have greater READING proficiency?		
Answer Options	Response Percent	Response Count
German	15.0%	6
English	65.0%	26
Same for both German and English	20.0%	8
		40

Figure 6. Reading proficiency language dominance for Group C

3.7 Grammaticality Judgments

Research techniques for accessing grammaticality and bilingual code-switching competence have been approached in linguistic literature in a variety of ways. While Sobin (1984), Dussias (2003), and McAlister (2010) use written tokens to illicit judgments, Aguirre (1985) presents his informants with an auditory stimulus, in order to reduce prescriptive attitudes towards switching.

Although many code-switched strings are either clearly acceptable or unacceptable, "a significant number of sentences fall somewhere in between in a gray area of partial acceptability" (Sprouse, 2007, p. 118). Several response methods for capturing reliable information about the gradience of grammaticality are available. Among those are category scale methods, ratio scale methods such as magnitude estimations and line drawing. "A category scale is simply a sequence of response categories that are understood to be uniformly spaced along some underlying continuum" (Cowart, 1997, p. 70). Because the extent of the scale is up to the individual researcher, judgment notation marking levels of

41

grammaticality remains ambiguous and plagued with inconsistencies. For the

purpose of this study, Andrews' (1990, p. 203) 6-point judgment scale will be

used, and is characterized as follows.

✓ : Completely acceptable and natural
? : Acceptable, but perhaps somewhat unnatural
??: Doubtful, but perhaps acceptable
?*: Worse, but not totally unacceptable
* : Thoroughly unacceptable
**: Horrible

Cowart (1997) contends that providing the informant with a rich scale which

potentially exceeds the number of categories inherent to the grammar, "no

information is lost by providing excess categories" (p. 67). The magnitude of the

scale ought to strive for the highest level of measurement which can most

appropriately capture the phenomenon in question. In a study done by Bader and

Häussler (2010), binary grammaticality judgments were compared to more

elaborate magnitude estimation procedures, revealing that gradient and binary

judgments lead to highly similar results. Arguably, the continuum in

grammaticality is reflective of a progression or gradience in grammatical

knowledge and ought to be formalized by the theory of grammar (Keller, 2000).

The main section of the survey instrument consists of code-switched German-

English tokens used to elicit grammaticality judgments from the informants.

Group A and Group B were given a written version of the questionnaire which

may be found in Appendix E, while group C was asked to listen to an audio

version of identical statements. The tokens were constructed based on data from a

naturalistic corpus study conducted by Eppler (2004) to counter criticism of

ecological validity. The 24 CS statements explored word order variation in

German and English dependent clauses following either *because* or *weil.* The

following table summarizes the distribution of the token set.

	Verb	*because*	*weil*
parE-depE (EGE2)	S		I don't eat meat, weil I a vegetarian **am**.
parG-depE (GGE2)	S		Ich esse kein Fleish, weil I a vegetarian **am**.
parE-depG (EEG2)	S	I don't eat meat because ich Vegetarier **bin**.	
parG-depG (GEG2)	S	Ich esse kein Fleisch because ich Vegetarier **bin**.	
parE-depE (EGE1)	M		I don't eat meat, weil I **am** a vegetarian.
parG-depE (GGE1)	M		Ich esse kein Fleisch, weil I **am** a vegetarian.
parE-depG (EEG1)	M	I don't eat meat, because ich **bin** Vegetarier.	
parG-depG (GEG1)	M	Ich esse kein Fleisch because ich **bin** Vegetarier.	

Table 2. Language of parent and dependent of because and weil, and position of
verb in dependent clause (S=subordinated or M=main)

The CS strings explore the variability of word order in dependent clauses headed

by *weil* or *because* with respect to finite verbal placement. The token set included

3 semantically distinct strings for each type of sentence, of which there were 8,

for a total of 24 tested tokens, which may be found in Appendix E.

Chapter 4

FINDINGS

This chapter presents the results from the German-English grammaticality judgment questionnaire. Data comparing the stability of written judgment data from Group B1 and Group B2 are presented first, followed by a comparison between the Group A judgment data elicited through written tokens and the Group C spoken token judgments. This is then followed by an analysis of differences in judgments given by German-dominant, English-dominant, and balanced bilinguals. The data for each analysis is presented first in terms of descriptive statistical methods, and then followed by inferential statistical analyses, using a correlated groups *t* test, and independent groups *t* test, or a one-way between-subjects analysis of variance, where appropriate.

4.1 Correlated Groups *t* Test

In order to determine whether the mean judgment scores from two groups are *statistically* different from one another, a Wilcoxon signed-rank test may be used in place of the correlated groups *t*-test The Wilcoxon signed-rank test is a nonparametric version of the correlated groups *t*-test, the latter of which is implemented when the groups being measured meet all of the following parameters below.

1. The scale of the measurement for both groups needs to have the properties of an equal-interval scale.

2. The difference between the paired values of both groups has been randomly drawn from the source population.

3. The source population from which these differences have been drawn can be reasonably supposed to have a normal distribution.

4. The groups being measured, and the testing measure being used must be identical.

An alternative to the matched-pair *t*-test, the Wilcoxon signed-sank test, which is an analogous alternative to the *t*-test for correlated sample, may be applied when the above mentioned assumptions are not met. Based on these guidelines, the matched-pair *t*-test is an appropriate measure for determining whether Groups B1 and B2 are statistically different from one another, given that the quantitative variables are measured on an ordinal level (from 1-6). The nonparametric Wilcoxon signed-rank test, however, makes fewer distributional assumptions than its parametric counterpart, and can be used to analyze quantitative variables which are measured on an ordinal level. The Wilcoxon signed-rank test compares correlated or identical groups in terms of differences between distributions of scores. As such, the Wilcoxon signed-rank test is an appropriate measure to detect differences between distributions of scores in Groups B1 and Groups B2, which look at the stability of grammaticality judgments for written tokens over time.

4.2 Independent Groups *t* Test

The independent groups *t* test is a statistical technique developed to analyze the relationship between two variables under the following conditions. The Mann-

Whitney *U* Test and Wilcoxon rank sum test are the nonparametric counterparts of the independent groups *t* test and may be used if one of the following parameters is not met, however, at the sacrifice of some power of the calculation.

1. The dependent variable is quantitative in nature and is measured on a level that at least approximates interval characteristics.

2. The independent variables are between subjects in nature.

3. The independent variable has two and only two levels.

The only difference between the Mann-Whitney *U* test and the Wilcoxon rank sum test is the computational procedure; however, both tests yield identical results. Additionally, the only difference between the Mann-Whitney *U* test and the Wilcoxon signed-rank test discussed in the previous section is that they are used for independent and correlated groups respectively. As such, either the independent groups *t* test, the Mann-Whitney *U* test or the Wilcoxon rank sum test are an appropriate measure for computing the relationship between Group A and Group C and determining whether written tokens are judged in a statistically different way than spoken versions of identical tokens.

4.3 ANOVA Measure

The one-way between-subjects analysis of variance (ANOVA) is frequently used to determine the relationship between variables when the following parameters are met.

1. The dependent variable is quantitative in nature and is measured on a level that at least approximates interval levels.

2. The independent variable is between-subjects in nature.

3. The independent variable has three or more levels.

Accounting for variance in CS within a group addresses a pertinent methodological issue in research on bilingualism, that is to say, that not all types of speakers are likely to switch in the same way or give the similar judgments for identical tokens. "An ANOVA compares the means of each group to the grand mean of all groups and then calculates deviations from the grand mean in order to establish within-groups and between-groups variation" (McAlister, 2010, p. 129).

Given the above mentioned parameters, the ANOVA is an appropriate measure for determining the statistical relationship between German-dominant, English-dominant, and balanced bilinguals' judgment data for Groups A, Groups C, and Groups A and C combined. Groups B1 and B2 are not considered for computation because Groups B1 and B2 are already subsets of Group A. In addition, a Tukey post hoc test for differences in ANOVA should be run, which "discerns the nature of the relationship by testing a null hypothesis for each possible pair of group means" (Jaccard & Becker, 1997, p. 340).

4.4 Stability of Written Tokens over Time

When testing for the stability of the written tokens over time, a correlated groups t test is the measure of choice. The hypotheses for the correlated groups t test are based on the first research question, which is given below.

1. Are German-English bilinguals' responses to individual code-switched written sentences stable over time?

H_0: The grammaticality judgments of German-English bilinguals will not differ over time.

H_1: The grammaticality judgments of German-English bilinguals will differ over time.

The independent variables are Group B1 and Group B2. The dependent variables include the composite response to each traditional written grammaticality judgment task, which is given in Table 3 below. The composite responses to the written grammaticality judgments are derived by computing the mean response for each token type. Figure 7 presents graphically the means of the data in Table 3.

		EEG1	EEG2	GEG1	GEG2	EGE1	EGE2	GGE1	GGE2	ALL
B1	M	2.97	3.96	3.27	3.80	3.46	5.21	3.94	4.56	3.90
	SD	1.56	1.62	1.36	1.48	1.29	0.99	1.55	1.36	1.58
B2	M	2.92	3.76	3.41	3.58	3.39	5.07	3.78	4.27	3.78
	SD	1.56	1.53	1.50	1.63	1.59	0.97	1.64	1.45	1.62

Table 3. Mean and standard deviation by switch type for Group B1 and Group B2
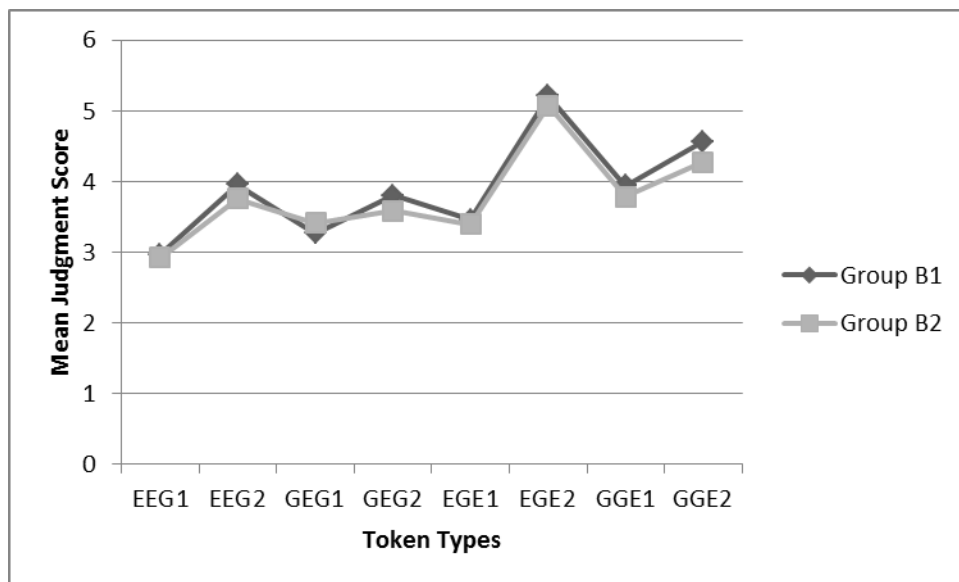

Figure 7. Mean score by token type for Group B1 and Group B2

It appears that token type EEG1, exemplified by sentences (29)-(31) below, is most preferred by both groups B1 and B2 with a mean rating of 2.97 and 2.92 respectively.

    (29)   I'm broke because meine Katze **trinkt** viel Milch.

                I am broke because my   cat   drinks a lot milk

                'I'm broke because my cat drinks a lot of milk.'

    (30)   I got a good grade because ich **habe** stundenlang **gelernt**.

                I got a good grade because I   have hour-long   learned

                'I got a good grade because I studied for hours.'

    (31)   I don't eat meat because ich **bin** Vegetarier.

                I don't eat meat because I   am vegetarian

                'I don't eat meat because I am a vegetarian.'

Alternatively, the least preferred switch in writing is token type EGE2 as seen in strings (32)-(34), with a mean rating of 4.56 given by Group B1 and 4.27 by Group B2.

    (32)   I'm broke, weil   my cat a lot of milk **drinks**.

                I am broke because   my cat a lot of milk drinks

                'I'm broke because my cat drinks a lot of milk.'

    (33)   I got a good grade,   weil   I for hours **studied**.

                I got a good grade   because I for hours studied

                'I got a good grade because I studied for hours.'

(34)   I don't eat meat, weil   I a vegetarian **am**.

        I don't eat meat because   I a vegetarian am

        'I don't eat meat because I am a vegetarian.'

Generally there appears to be no significant change in written judgments between groups over time. While composite judgment score for Group B1 for all tokens is 3.9, Group B2 gave a slightly more acceptable score of 3.78. The lack of a robust difference in judgment scores between Groups B1 and B2 is confirmed by the correlated groups *t* test. For an alpha level of .05, the rank sums (N=34) were found not to differ significantly with a significance level of .271, suggesting that written code-switched tokens are judged similarly by Group B1 (M=3.90, SD=.91) as they are by Group B2 (M=3.78, SD=1.12). As such, the null hypothesis is retained, indicating that German-English bilinguals' judgments to written tokes are stable over time.

In like manner, the correlated pairs *t* test was run for the composite score of each token type. A Bonferroni correction was applied to correct the problem of multiple comparisons, yielding an alpha level of .0625. While token types EEG1, EEG2, GEG1, GEG2, EGE1, EGE2, and GGE1 were found to have a significance level of .208 or higher, retaining the null hypothesis, token type GGE2 (35)-(37) exhibited a significance level of .020, indicating that for this type of switch, the judgments may not be stable over time.

(35)   Ich habe kein Geld,   weil   my cat a lot of milk **drinks**.

        I   have no   money   because   my cat a lot of milk drinks

        'I don't have any money because my cat drinks a lot of milk.'

50

(36)   Ich habe eine gute Note bekommen, weil   I for hours **studied**.

         I   have a   good grade received   because I for hours studied

         'I got a good grade because I studied for hours.'

(37)   Ich esse kein Fleisch, weil   I a vegetarian **am**.

         I   eat no   meat   because   I a vegetarian am

         'I don't eat meat because I'm a vegetarian.'

Even though Group B1 and Group B2 gave consistent judgments between groups as indicated by the correlated groups *t* test, the informants in each group did not respond uniformly. In addition to the statistically significant discrepancy for tokens (35)-(37) across groups, both groups exhibited informant discrepancy within each respective group as indicated by the standard deviations.

4.5 Comparison of Written and Spoken Tokens

For the comparison between independent variables Group A and Group C, an independent groups *t* test was performed. The hypotheses for the independent groups *t* test are based on the second research question, repeated below.

2. Do judgments of German-English bilinguals to individual code-switched sentences differ when the tokens are presented visually in contrast to being presented aurally?

    H_0: The grammaticality judgments of German-English bilinguals will not differ between written and spoken tokens.

    H_1: The grammaticality judgments of German-English bilinguals will differ between written and spoken tokens.

51

The dependent variables include the composite response to each traditional written and spoken grammaticality judgment task, which is given in Table 4 below. The composite responses to the written or spoken grammaticality judgments are derived by computing the mean response for each token type. Figure 8 presents graphically the means of the data in Table 4.

Table 4 – Average by switch type for Group A and Group C										
		EEG1	EEG2	GEG1	GEG2	EGE1	EGE2	GGE1	GGE2	ALL
A	Mean	2.88	3.84	3.32	3.76	3.53	5.06	3.87	4.41	3.83
	StDev	1.61	1.58	1.49	1.48	1.44	1.08	1.56	1.50	1.61
C	Mean	2.78	3.15	3.03	3.14	3.36	4.98	3.73	4.27	3.56
	StDev	1.35	1.39	1.47	1.56	1.51	1.04	1.34	1.37	1.55

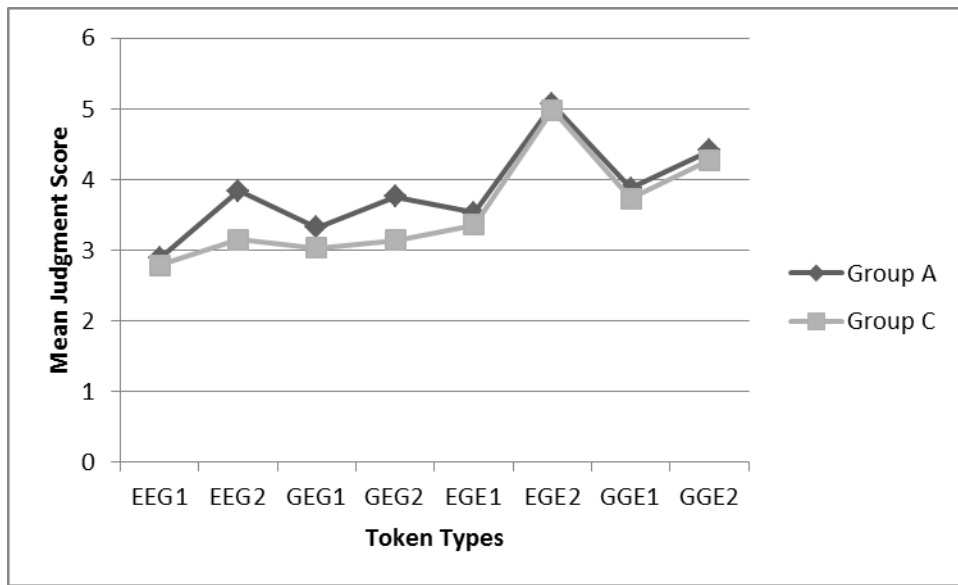*Table 4. Mean and standard deviation by switch type for Group A and Group C*

Figure 8. Mean score by token type for Group B1 and Group B2

It appears that token type EEG1, exemplified by sentences (29)-(31) in the previous section, is most preferred by both groups A and C with a mean rating of 2.88 and 2.78 respectively. Alternatively, the least preferred switch in writing is token type EGE2 as seen in tokens (32)-(34), with a mean rating of 5.06 given by Group A and 4.98 by Group B2.

52

Generally there appears to be no significant difference between written and spoken judgment tasks. While the composite judgment score for Group A for all tokens is 3.83, Group C gave a slightly more acceptable score of 3.56. The lack of a robust difference in judgment scores between groups A (N=50) and C (N=40) is confirmed by the independent groups t test comparing the mean judgment scores for written tokens (M=3.83, SD= 1.04) and the spoken tokens (M=3.56, SD=.83). For an alpha level of .05, this test was found to be statistically insignificant with a significance level of .186 assuming equal variance and .175 not assuming equal variance, suggesting that written code-switched tokens are judged similarly to spoken tokens by German-English bilinguals. This results in the verification of the null hypothesis, indicating that German-English bilinguals' judgments will not differ between written and spoken tokens.

Similarly, the independent groups t test was run for the composite score of each token type. A Bonferroni correction was applied to correct the problem of multiple comparisons, yielding an alpha level of .0625. Token types EEG1, GEG1, EGE1, EGE2, GGE1, and GGE2 were found to have significance level of .292 or higher, retaining the null hypothesis. For token type EEG2 (38)-(40) and GEG2 (41)-(43), however, the test was shown to be statistically significant with values of .013 (equal variance assumed) and .033 (equal variance assumed) respectively, indicating that for those two token types, judgments vary when the informant is presented with a written or a spoken token.

(38)   I'm broke because meine Katze viel   Milch **trinkt**.

       I am broke because my   cat   much   milk drinks

       'I'm broke because my cat drinks a lot of milk.'

(39)   I got a good grade because ich stundenlang **gelernt habe**.

       I got a good grade because I   hour-long   studied have

       'I got a good grade because I studied for hours.'

(40)   I don't eat meat because ich Vegetarier **bin**.

       I don't eat meat because I   vegetarian am

       'I don't eat meat because I am a vegetarian.'

(41)   Ich habe kein Geld because   meine Katze viel   Milch **trinkt**.

       I   have no   money because my   cat   much   milk drinks

       'I'm broke because my cat drinks a lot of milk.'

(42)   Ich habe eine gute Note bekommen because ich stundenlang

       I   have a   good grade received   because I   hour-long

       gelernt habe.

       studied have

       'I got a good grade because I studied for hours.'

(43)   Ich esse kein Fleisch because ich Vegetarier **bin**.

       I   eat no   meat   because I vegetarian   am

       'I don't eat meat because I am a vegetarian.'

Despite the general agreement in judgments between groups as indicated by the

independent groups *t* test, Group A and Group C did not always respond

uniformly. In addition to the statistically significant discrepancy for tokens (38)-

(43) across groups, both groups exhibited informant discrepancy within each respective group as indicated by the standard deviations.

4.6 Comparisons of Judgments between Types of Bilinguals

For the comparison of judgment between types of bilinguals, namely German-dominant, English-dominant, and balanced, the ANOVA measure was performed. The hypotheses for the ANOVA test are based on the third research question.

3. Do German-English bilinguals evidence differences in grammaticality judgments as a function of language dominance?

    H_0: German-English bilinguals do not evidence differences in grammaticality judgments as a function of language dominance.

    H_1: German-English bilinguals do evidence differences in grammaticality judgments as a function of language dominance.

The dependent variables, which include the composite response to each traditional written and spoken grammaticality judgment task, are given for Group A in Table 5 below. Figure 9 presents graphically the means of the data in Table 5. The composite responses to the written or spoken grammaticality judgments are derived by computing the mean response for each token type. The independent variables are the types of bilinguals, either German-dominant, English-dominant, or balanced bilinguals, in Groups A, Groups C, and Groups A and C together.

Group A (N=50)		EEG1	EEG2	GEG1	GEG2	EGE1	EGE2	GGE1	GGE2	ALL
G	M	3.70	5.27	3.64	5.09	3.79	5.39	4.58	4.80	4.51
(N=11)	SD	1.75	0.94	1.62	0.94	1.47	0.86	1.60	1.64	1.56
E	M	2.71	3.39	3.25	3.33	3.48	4.87	3.67	4.28	3.61
(N=34)	SD	1.56	1.50	1.51	1.40	1.46	1.15	1.52	1.46	1.59
B	M	2.4	3.73	3.07	3.8	3.27	5.6	3.73	5	3.82
(N=5)	SD	0.94	1.43	0.96	1.32	1.22	0.51	1.39	1.21	1.50

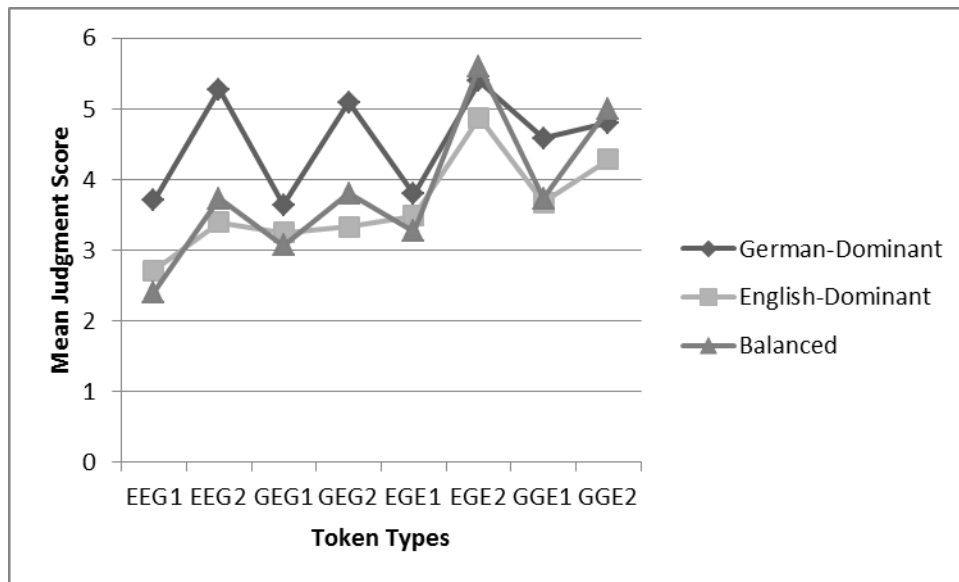*Table 5. Mean and standard deviation by switch type by language dominance for Group A*

Figure 9. Mean score by token type for language dominance for Group A

It appears that token type EEG1, exemplified by sentences (29-31), is most preferred by both English-dominant (N=34) and balanced bilinguals (N=5) with a mean rating of 2.71 and 2.40 respectively. For German-dominant bilinguals (N=11) GEG1 is most preferred with a mean rating of 3.64, however token type EEG1 does not differ significantly with a mean score of 3.70. Alternatively, the least preferred switch in writing is token type EGE2, as discussed in the previous two sections (32-35). German-dominant, English-dominant, and balanced bilinguals assigned mean scores of 5.39, 4.87, and 5.6 respectively. There appears

to be quite a significant difference between the judgments given by different types of bilinguals. While the composite judgment score for the German-dominant informants for all written tokens is 4.87, the English-dominant and balanced bilinguals gave a slightly higher mean rating of 3.61 and 3.82. Based on the graphical representation of the data, it is evident that the English-dominant and balanced bilinguals elicit scores which map more closely onto one another than their German dominant counterparts.

The robust difference in mean judgment scores between types of bilinguals is confirmed by the ANOVA test. The alpha level was 0.05. There was a statistically significant difference of .041 at the $p<.01$ level for the three groups The effect size, calculated using eta squared, was .127. Post-hoc comparisons using the Tukey HSD test indicated that the normed written token mean for German-dominant bilinguals (M=4.51, SD=1.56) was significantly different from the English-dominant bilinguals (M=3.61, SD=1.56). The other groups did not differ significantly from one another.

A one-way between-groups analysis of variance was also conducted in order to compare the difference in German-dominant (N=6), English-dominant (N=29), and balanced bilinguals (N=5) composite judgment responses for the spoken tokens presented to Group C, which are presented in Table 6 below and represented graphically in Figure 10.

Group C (N=40)		EEG1	EEG2	GEG1	GEG2	EGE1	EGE2	GGE1	GGE2	ALL
G	M	2	3.83	2.72	3.89	2.44	4.56	3.06	3.67	3.24
(N=6)	SD	0.88	1.15	1.41	1.28	1.34	1.50	1.51	1.61	1.55
E	M	2.94	3.01	3.17	3.01	3.66	5.20	3.86	4.54	3.68
(N=29)	SD	1.37	1.39	1.50	1.57	1.51	0.86	1.27	1.31	1.56
B	M	2.75	3.13	2.6	3	2.73	4.20	3.73	3.90	3.26
(N=5)	SD	1.41	1.55	1.30	1.65	1.10	0.86	1.39	1.12	1.40

Table 6. Mean and standard deviation by switch type by language dominance for Group C
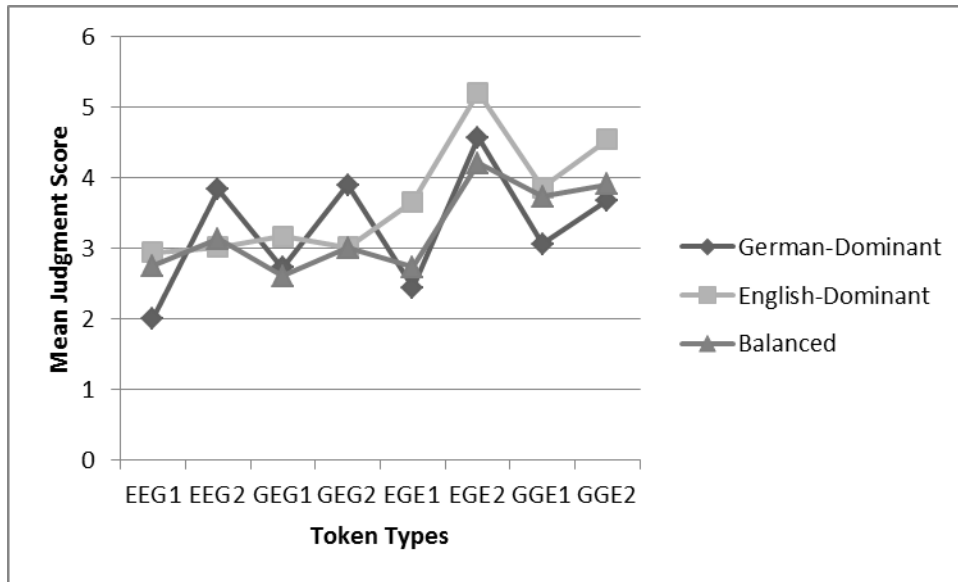

Figure 10. Mean score by token type for language dominance for Group C

For Group C, the lack of a statistically robust difference in mean judgment scores between types of bilinguals is confirmed by the ANOVA test. The alpha level was 0.05. There was no statistically significant difference at the $p<.01$ level for the three groups where p= .35. The effect size, calculated using eta squared, was .055. Post-hoc comparisons using the Tukey HSD test indicated that the normed spoken token mean for German-dominant bilinguals (M=3.24, SD=1.55) did not significantly differ from the English-dominant bilinguals (M=3.68, SD=1.56) or the balanced bilinguals (M=3.26, SD=1.40).

Group A and C (N=90)		EEG1	EEG2	GEG1	GEG2	EGE1	EGE2	GGE1	GGE2	ALL
G	M	3.10	4.76	3.31	4.67	3.31	5.10	4.04	4.40	4.06
N=17	SD	1.70	1.23	1.59	1.19	1.56	1.19	1.72	1.70	1.67
E	M	2.81	3.22	3.21	3.19	3.56	5.02	3.76	4.40	3.64
N=63	SD	1.48	1.46	1.50	1.48	1.48	1.04	1.41	1.40	1.57
B	M	2.58	3.43	2.83	3.40	3.00	4.90	3.73	4.45	3.54
N=10	SD	1.20	1.50	1.15	1.52	1.17	0.99	1.36	1.28	1.48

Table 7. Mean and standard deviation by switch type by language dominance for Group A and Group C combined
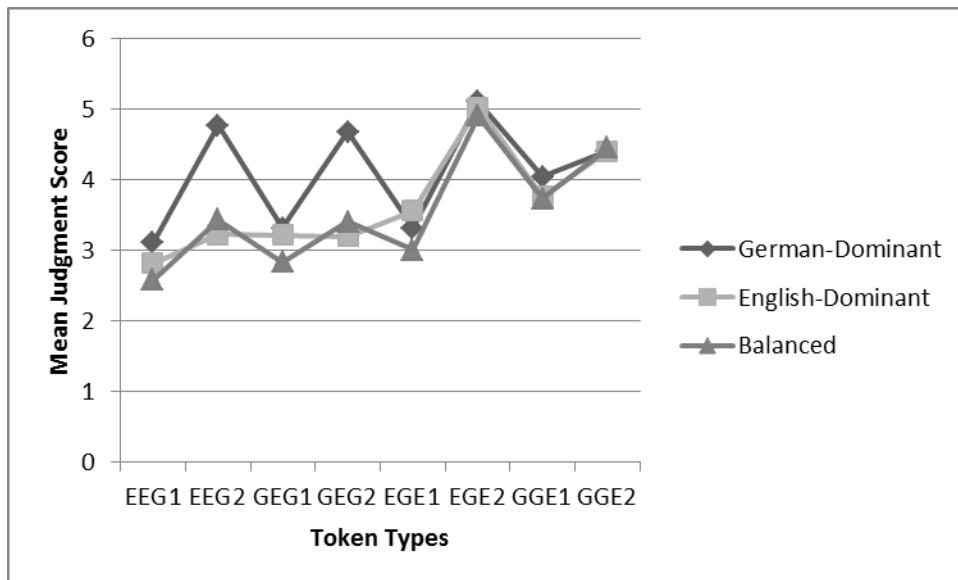

Figure 11 Mean score by token type for language dominance for Group A and Group C combined

When groups A and C are combined, the lack of a statistically robust difference in mean judgment scores between types of bilinguals is confirmed by the ANOVA test. The alpha level was 0.05. There was no statistically significant difference at the $p<.01$ level for the three groups, where p= .234. The effect size, calculated using eta squared, was .033. Post-hoc comparisons using the Tukey HSD test indicated that the normed written and spoken token mean for German-dominant bilinguals (M=4.06, SD=1.67) did not significantly differ from the

English-dominant bilinguals (M=3.64, SD=1.57) or the balanced bilinguals (M=3.54, SD=1.48).

Although the ANOVA measure confirmed that there was general agreement between the different bilingual groups between the composite mean scores of all tokens, the bilingual groups did not respond uniformly for all tokens if each token type were to be examined separately. The judgment pattern between German-dominant, English-dominant, and balanced bilinguals remained generally consistent, meaning that all groups gave a less acceptable rating to token types EEG1 than to EEG2. Similarly, all bilingual types considered EGE2 as the least acceptable token. Disagreement between groups was more prominent in certain syntactic contexts.

Results from this study indicate that there are some differences in the judgments of German-English bilinguals for code-switched *weil* and *because* clauses. Although informants remained consistent with their judgments of spoken tokens over time, string types EEG2 and GEG2 showed the greatest amount of disagreement between groups A and C, which were presented with written and spoken tokens respectively. Group C gave a higher normed token mean for those particular constructions. Similarly, when comparing German-dominant, English-dominant, and balanced bilinguals, token types EEG2 and GEG2 were judged significantly less acceptable by the German-dominant bilinguals than by the other two groups. This remains a consistent pattern between all 90 informants. This perhaps indicates that balanced and English-dominant bilinguals are more prone to code-switching. This assumption should be considered delicately, given that the

distribution of bilingual types was disproportionate; while 70% of all participants

were English-dominant, only 18% were German-dominant, and 12% were

balanced bilinguals.

Chapter 5

CONCLUSION

The main purpose of this thesis is to explore the stability of grammaticality judgments given for written tokens by German-English bilinguals, as formulated by the primary research question below.

Are German-English bilinguals' responses to individual code-switched written sentences stable over time?

In addition, the study investigates differences in CS patterns for identical tokens of different modalities, namely written in contrast to spoken. This aim is reflected in the secondary research question, which is repeated here.

Do judgments of German-English bilinguals to individual code-switched sentences differ when the tokens are presented in writing in contrast to being spoken?

The third research question

Do German-English bilinguals evidence differences in grammaticality judgments as a function of language dominance?

examines differences in judgments which may result as a consequence of language dominance.

5.1 Elicited Data for Written Tokens Judged over Time

The German-English grammaticality judgment data, which was outlined in Chapter Four, indicates that German-English bilingual study participants (N=34) gave almost identical judgment responses for code-switched tokens when

presented with the written sentence the first time, followed by the second time a few weeks later. The results from this particular study stand in direct contrast and opposition to work done by Ellis (1991), Davies and Kaplan (1998), Sorace (1988), and support the reliability of grammaticality judgment data as has been explicitly argued by Gass (1994), Leow (1996), and Mandell (1999). Ellis, who found that the advanced Chinese learners of English he used as participants were inconsistent in a considerable portion of their judgments as a result of uncertainty, mirrors Sorace's findings who states that the majority of reliability issues are a result of indeterminacy of the tokens and "the absence of a clear grammaticality status for particular linguistic constructions in the speaker's competence" (1988, p. 166). The grammaticality judgments in this particular study, however, remain stable over time, even though participants gave ratings which were indicative of linguistic indeterminacy. Chaudron (1983), who reviewed 39 studies investigating the stability and reliability of grammaticality judgment data, concluded that "given appropriate controls and validation procedures, metalinguistic judgments can play a useful role in language acquisition studies" (p. 343). As a result, it will be concluded that grammaticality judgments are a viable method for studying German-English CS from a structure based approach, given that the methods utilized for the collection of data are appropriately controlled.

5.2 Elicited Data for Written versus Spoken Tokens

A similar conclusion may be given when comparing written tokens to spoken tokens. While Birdsong (1989) suggests that the problem of the stability

of grammaticality judgment data is particularly acute when the strings being tested occur in speech but are proscribed in writing, as is the case for CS, the data from this study indicates that bilingual informants from two different groups judge tokens in a consistent manner across groups, regardless of whether the token was presented in writing or speech. On account of the findings of this study, it will be concluded that for the particular tokens used for this study judged by highly proficient German-English bilinguals, the modality of the tokens will elicit a slight variation, but nothing that is statistically significant and cannot be attributed to other variables and factors, and as such as suggested by these findings may be used interchangeably for data collection.

5.3 Elicited Data by Language Dominance

Data from this particular study show that judgments from German-dominant, English-dominant, and balanced bilinguals generally do not vary in a statistically significant way when the composite scores from all token types are compared via an ANOVA measure. The data do indicate however, that different types of bilinguals varied their acceptance of certain switches in different syntactic contexts, most notably for token types EEG2 and GEG2, in which the dependent clause was German and subordinated. For these particular string types, the German-dependent bilinguals gave statistically significantly less acceptable ratings than their English-dominant and balanced bilingual counterparts. It ought to be noted, however, that not all German-dominant participants preferred a lower rating for token types EEG2 and GEG2. In a similar manner, while the English-

dominant and balanced bilinguals generally gave higher ratings of acceptability, not all participants from those groups gave ratings in an identical manner. Some German-dominant bilinguals shared intuitions with the English-dominant and balanced bilingual groups, while some English-dominant and balanced bilinguals shared intuitions with the German-dominant group. Although the groups disagreed between themselves significantly for two token types, their judgments remain consistent for all other token types, indicating that the participants shared the same intuitions in those instances regardless of language dominance. It appears that language dominance in bilinguals may affect intuitions about grammaticality; however, this effect is not felt for all bilingual types and syntactic contexts. Given the uneven distribution of German-dominant and balanced bilinguals to English-dominant bilingual, this particular phenomenon ought to be retested.

5.4 Implications for Research Methods in Code-switching

While research to date on the reliability of grammaticality judgment data has yielded mixed results, "it is clear that the incongruence stemmed, at least in part, from a difference in the approach adopted" (Han, 2006, p. 61). Although some prominent researchers in CS, such as Myers-Scotton (2006), argue for a higher fidelity of naturalistic data than elicited data, this study shows that traditional grammaticality judgments are sufficiently stable over time. When grammaticality judgments are analyzed quantitatively and validated though other comparative measures, such as they were done in this particular study, rather than

qualitatively, "evidence supporting the reliability of grammaticality judgment tasks by researchers" (Han, 2006, p. 61) has been made available. Given that the tokens for this study were chosen based on a small corpus of naturalistic German-English data set, the concern over the ecological validity of the tokens is no longer valid. On a similar note, based on the corpus used by Eppler 2004, only token types GEG1, GEG2, and EEG1 are elicited and recorded. The other 5 token types used in this study to test the position of the verb in the code-switched dependent clause, would not be expected to be allowed by the grammars of the two languages. Eppler's corpus data is largely supported by the grammaticality judgment data from this study. While token type EEG1 was given the highest mean composite score of 2.83, GEG1 was second with 3.12. However, token type EGE1 received a mean composite score of 3.29, also significantly lower than the rest of the strings. Its absence in the corpus may be explained through the homophone-like quality of the conjunction *weil* to the English *while*, and could be strategically avoided in discourse by German-English bilinguals in order to avoid adding unnecessary ambiguity.

The data from the German-English grammaticality judgment tasks indicates that there are differences among German-dominant, English-dominant, and balanced bilinguals. The differences, however, do not necessarily implicate the exclusion of language dominant bilinguals from CS studies.

5.5 Limitations and Further Research

Throughout this thesis I have attempted to illustrate how linguistics, with special emphasis on the experimental syntax of code-switching, can be treated as an empirical science. The approaches which were used in designing the experiment, collecting the data, and evaluating the results were heavily influenced by various quantitative, rather than qualitative, methods and approaches. Despite aiming to be as empirically sound as possible, certain elements could not be controlled for in this study.

While the participants were carefully selected, it is impossible to find a homogenous group of monolingual speakers, on account of dialectal differences, individual speaker preferences, sociolinguistic variables, physiological variation, and so forth. The problem of homogeneity is only compounded when considering bilinguals. In her dissertation, McAlister (2010) discusses aspects unique to research construction in bilingualism, given that bilingual language production and assessment pose additional variables which ought to be considered by the researcher. She discovered that while early sequential, late sequential and simultaneous bilinguals gave different judgments, the discrepancy was not significant enough to warrant the exclusion of one group over another. In order to eliminate the amount of unknown variables, participants for this study were carefully selected based on prior research. As a result, age of acquisition was not considered. Similarly, no formal language assessment tool was utilized, mainly because no good measure exists which would be compatible across languages. Work by Montrul (2009) suggests that a speaker's self-assessment is stable and

accurate enough to be considered for empirically based studies. Although the studies by Montrul and McAlister were intended to be extended and applied to other language pairs, they were tested on Spanish-English and Sovak-English bilinguals. As such, the assumptions made for this study, based on research by McAlister and Montrul, may or may not be valid.

Grammaticality judgment tasks, at least traditionally, are a fantastic tool for testing and discovering the syntactic structure or neutral framework of a language. The tokens for this study were based on tokens found in naturalistic data, so that no question of ecological validity could arise. In addition, the tokens were designed on a semantic level to be as culturally generic as possible, in order to remain relatable and understandable to all participants. However, the "coupling between word meanings and innovative human thought means that word meanings have an unpredictability that, arguably, makes them incapable of being brought within the purview of empirical scientific theorizing" (Sampson, 2001, p. 184). When we study the syntax of natural languages, it is impossible to ignore the semantic aspect, which Sampson (2001) considers a foremost problem to the scientific study of language. While analyzing the data from this study, it became apparent that one of the three tokens constructed for each switch type was slightly less preferred than the other two, regardless of syntactic context. Given that this was consistent across all switch types, the lower acceptability could be on account of some phonological, morphological, or semantic interference for the participants. The semantic-syntax interface, as such, should be considered

carefully, perhaps followed by a formal study of grammaticality judgment trends for code-switched strings of varying levels of semantic acceptability.

In conclusion, this thesis has provided evidence and demonstrated that written grammaticality judgments of various code-switched tokens remain stable over time. Similarly, the study suggests that there is no measurable difference between the judgment of written and spoken tokens. When comparing differences in grammaticality judgments by language dominance, there was some evidence for differences in judgments, however, it is not clear whether the difference is as a result of language dominance or other external factors. As a result, grammaticality judgment tasks are a viable and robust research tool for syntactic aspects of code-switching, given that certain methodological practices, such as the participants involved in the study, the instructions, tokens, sample size, and judgment scale are carefully controlled. I hope that this thesis has aided in quieting some of the criticism surrounding grammaticality judgments for the collecting of linguistic evidence, and has expounded on how we can continue to produce empirically sound and replicable research.

REFERENCES

Adli, A. (2005). Gradedness and Consistency in Grammaticality Judgments. In S. Kepser & M. Reis (Eds.), *Linguistic evidence: Empirical, theoretical, and computational perspectives* (pp. 7-25). Berlin: Walter de Gruyter.

Aguirre, A. (1977). *Acceptability judgment of grammatical and ungrammatical forms of intrasentential code alteration.* Doctoral dissertation, Stanford University, Stanford, CA.

Aguirre, A. (1985). An experimental study of code alternation. *International Journal of the Sociology of Language*, 53, 59-81.

Andrews, A. (1990). Case structures and control in Modern Icelandic. In J. Maling & A. Zaenen (Eds.), *Modern Icelandic syntax* (pp. 187-234). Syntax and Semantics 24. San Diego: Academic Press.

Auer, P. (1998). *Code-switching in conversation: Language, interaction and identity*. London: Routledge.

Bader, M., & Häussler, J. (2010). Toward a model of grammaticality judgments. *Journal of Linguistics*, 46(2), 273-330.

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

Belazi, H.M. (1991). Multilingualism in Tunisia and French/Arabic codeswitching amonge ducated Tunisian bilinguals. Ithaca, NY. Cornell University dissertation.

Belazi, H.M., Rubin, E.J., & Toribio, A.J. (1994). Code switching and X-bar theory: The functional head constraint. *Linguistic Inquiry*, 25(2), 221-237.

Bentahila, A., & Davies, E. (1983). The syntax of Arabic-French code-switching. *Lingua*, 59(4), 301-330.

Bever, T.G. (1970a). The Cognitive basis for Linguistic Structures. In R. Hayes (Ed.), *Cognition and language development*, (pp. 279-362). New York:Wiley & Sons.

Bever, T.G. (1970b). The influence of speech performance on linguistic structures. In G. B. Flores d'Arcais and W. J. M. Levelt (Eds.), *Advances in Psycholinguistics*, (pp. 4-30). Amsterdam: North-Holland.

Bever, T.G., & Langendoen, T. (1971). A dynamic model of the evolution of language. *Linguistic Inquiry*, 2, 433-463.

Bever, T. G. (1974). The ascent of the specious, or there's a lot we don't know about mirrors. In D. Cohen (Ed.), *Explaining Linguistic Phenomena* (pp. 173-200). Washington: Hemisphere.

Bhatia, T.K., & Ritchie, W.C. (1996). Bilingual language mixing, universal grammar, and second language acquisition. In W.C. Ritchie & T.K. Bhatia (Eds.), *Handbook of second language acquisition*, (pp. 627-688). New York: Academic Press.

Birdsong, D. (1989). *Metalinguistic performance and interlinguistic competence*. Berlin: Springer-Verlag.

Bock, J. (1986). Syntactic persistence in language production. *Cognitive Psychology,* 18(3), 355-387.

Boztepe, E. (2003). Issues in code-switching: competing theories and models.*Working Papers in TESOL & Applied Linguistics* 3(2): 1-27.

Brown, R. (1973). A first language: The early stages. Cambridge, MA: Harvard University Press.

Bullock, B. E., & Toribio, A. J. (2009). *The Cambridge handbook of linguistic code-switching*. Cambridge: Cambridge University Press.

Bullock, B. E., & Toribio, A. J. (2009). Themes in the study of code-switching. In Bullock & Almeida, 2009, 1-17.

Cantone, K., & Müller, N. (2008). *Un nase or una nase*? What gender marking within switched DPs reveals about the architecture of the bilingual language faculty. *Linguia,* 118(6), 810-826.

Carol, J.M., Bever, T.G., & Pollack, C.R. (1981). The non-uniqueness of linguistic intuitions. *Language,* 57, 368-383.

Carstensen, B. (1965). *Englische Einflüsse auf die deutsche Sprache nach 1945*. Heidelberg: Universitätsverlag.

Carstensen, B., Viereck, W., & Bald, W.-D. (1986). *English in contact with other languages: Studies in honour of Broder Carstensen on the occasion of his 60th birthday*. Budapest: Akadémiai Kiadó.

Chaudron, C. (1983). Research in metalinguistic judgments: A review of theory, methods, and results. *Language Learning*, 33, 343-377.

Chomsky, N. (1964). *Current issue in linguistic theory*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the Theory of Syntax.* Cambridge, MA: MIT Press.

Chomsky, N. (1995). *The minimalist program.* Cambridge: MIT Press.

Chomsky, N., &Lasnik, H. (1977). Filters and control. *Linguistic Inquiry*, 8, 425-504.

Cowart, W. (1997). *Experimental syntax: Applying objective methods to sentence judgements*. Thousand Oaks, Calif: Sage Publications.

Crystal, D. (1997). *The Cambridge Encyclopedia of Language* (2nd ed.)*. Cambridge: Cambridge University Press.

Davies, W., & Kaplan, T. (1998). Native speaker vs. L2 learner grammaticality judgments. *Applied Linguistics*, 19(2), 183-203.

DiSciullo, A., Muysken, P., & Singh, R. (1986). Government and code-mixing. *Journal of Linguistics*, 22(1), 1-24.

Dussias, P. E. (2003). Spanish-English code-mixing at the auxiliary phrase: Evidence from eye movement data. *Revista Internacional de Lingüística Iberoamericana*, 2, 7-34.

Ellis, R. (1991). Grammaticality judgments and second language acquisition. *Studies in Second Language Acquisition*, 13, 161-186.

Eppler, E. (1999). Word order in German-English mixed discourse. *UCL Working Papers in Linguistics* 11, 285–308.

Eppler, Eva. (2004). "…because dem Computer brauchst' es ja nicht zeigen.": because + German main clause word order. *International Journal of Bilingualism,* 8(2), 127-144.

Farrar, K. (1999). Explanation for word order change in modern German. *ZDL* 55.

Han, Z. (2006). Fossilization: Can grammaticality judgment be a reliable source of evidence? In Z. Han, & T. Odlin (Eds.), *Studiesof Fossilization in Second Language Acquisition* (pp. 56-82). Clevedon: Multilingual Matters

Gardner-Chloros, P. (2009) Sociolinguistic factors in code-switching. In Bullock & Almeida, 2009, 97-114.

Gass, S. (1994). The reliability of second-language grammaticality judgments. In E. Tarone, S. Gass & A. Cohen (Eds.), *Research Methodology in Second-Language Acquisition* (pp. 303-322). Hillsdale, NJ: Lawrence Erlbaum Associates

Gleitman, H., & Gleitman, L. (1979). Language use and language judgment. In C. Fillmore, D. Kempler & W. Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 103-126). New York: Academic Press.

Goodluck, H. (1991). *Language acquisition: A linguistic introduction.* Cambridge: Blackwell.

Grimshaw, J., & Rosen, S. (1990). Knowledge and obedience: The developmental status of the binding theory. *Linguistic Inquiry*, 21, 187-222.

Gullberg, M., Indefrey, P., & Muysken, P. (2009). Research techniques for the study of code-switching. In Bullock, & Toribio, 2009, 21-39.

Gumperz, J. (1967). On the linguistic markers of bilingual communication. *Journal of Social Issues*, 28(2), 48-57.

Gumperz, J. (1970). Verbal strategies and multilingual communication. In J. E. Alatis (Ed.), *Report of the Twenty-first Annual Round-Table Meeting on Linguistics and Language Studies* (pp. 129-147). Washington, DC: Georgetown University Press.

Gumperz, J. (1976). The sociolinguistic significance of conversational code-switching. In J. Cook-Gumperz and J.J. Gumperz (Eds.), *Papers on language and context: Working Papers No.* 46, pp. 1-46. Berkeley, CA: University of California Language Behavior Research Laboratory.

Gumperz, J. (1982). *Discourse strategies.* Cambridge, UK and New York: Cambridge University Press.

Günther, S. (1996). From subordination to coordination? *Pragmatics*, 6, 323-356.

Günther, S. (1993). "…weil—man kann es ja wissenschaftlich untersuchen." *Linguistische Berichte*, 143, 37-59.

Hawkins, J. (1986). *A Comparative Typology of English and German.* London & Sidney: Croom Helm.

Jaccard, J., & Becker, M. A. (1997). *Statistics for the behavioral sciences*. Pacific Grove: Brooks/Cole.

Joshi, A. (1985). Processing of sentences with intrasentential code switching. In D.R. Dowty, L. Karttunen & A.M. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives* (pp. 190-205). Cambridge: Cambridge University Press.

Keller, F. (2000). Gradience in grammar: Experimental and computational aspects of degrees of grammaticality. Ph.D. dissertation, University of Edinburgh.

Kepser, S., & Reis, M. (2005). *Linguistic evidence*. Berlin: Mouton de Gruyter.

Küper, Christoph. (1991). Geht die Nebensatzstellung im Deutschen verloren? Zur pragmatischen Funktion der Wortstellung in Haupt- und Nebensätzen. *Deutsche Sprache*, 19, 133-158.

Labov, W. (1972a). *Sociolinguistic patterns.* Philadelphia: University of Pennsylvania Press.

Labov, W. (1972b). Some principles of linguistic methodology. *Language in Society*, 1, 97-120.

Labov, W. (1975). *What is a linguistic fact?* Lisse: Peter de Ridder.

Lehman, C. (2004). Data in linguistics. *The Linguistic Review*, 21, 175-210.

Lenth, R.V. (2001). Some practical guidelines for effective sample size determination. *American Statistician,* 55, 187-193.

Leow, R. (1996). Grammaticality judgment tasks and second-language development. *Georgetown University Round Table on Languages and Linguistics* (pp. 126-139). Washington, DC: Georgetown University Press.

Levelt, W., van Gent, J., Haans, A., & Meijers, A. (1977). Grammaticality, paraphrase, and imagery. In S. Greenbaum (Ed.), *Acceptability in language* (pp. 87-101). The Hague: Mouton.

Liceras, J., Fuertes, R., Perales, S., Pérez-Tattam, R., & Spradlin, K. (2008). Gender and gender agreement in bilingual native and non-native grammars: A view from child and adult functional lexical mixings. *Linguia,* 118(6), 827-851.

Maclay, H., & Sleator, M. (1960). Responses to language: Judgments of grammaticalness. *International Journal of American Linguistics*, 26, 275-282.

MacSwan, J. (1999). *A minimalist approach to intrasentential code switching*. Outstanding dissertations in linguistics. New York: Garland Pub.

MacSwan, J., & McAlister, K.T. (2010). Naturalistic and Elicited Data in Grammatical Studies of Codeswitching. *Studies in Hispanic and Lusophone linguistics, 3*(2), p.

Mahootian, S. (1993). A null theory of code switching. Ph.D. dissertation, Northwestern University.

Mandell, P.B. (1999). On the reliability of grammaticality judgment tests in second language acquisition research. *Second Language Research*, 15(1), 73-100.

McAlister, K. T. (2010). *Age of onset of exposure in codeswitching.* Ph.D. dissertation, Arizona State University.

McClure, E. (1981). Formal and functional aspects of the codeswitched discourse in bilingual children. In R.P. Durán (Ed.), *Latino language and communicative behavior* (pp. 69-94). Norwood, NJ: Ablex.

Meisel, J. (1994). Code-switching in young bilingual children: The acquisition of grammatical constraints. *Studies in Second Language Acquisition*, 16(4), 413-439.

Milroy, L., & Muysken, P. (1995). *One speaker, two languages: Cross-disciplinary perspectives on code-switching.* Cambridge, UK and New York: Cambridge University Press.

Muysken, P. (1995). Code-switching and grammatical theory. In L. Milroy & P. Muysken (Eds.), *One speaker two languages: Cross-disciplinary perspectives on code-switching* (pp. 177-198). New York: Cambridge University Press.

Muysken, P. (2000). *Bilingual speech: A typology of code-mixing.* Cambridge: Cambridge University Press.

Myers-Scotton, C. (1993). *Duelling languages: Grammatical structure in codeswitching.* Oxford and New York: Oxford Univerity Press.

Myers-Scotton, C. (2006). Natural codeswitching knocks on the laboratory door. *Bilingualism: Language and Cognition*, 9(2), 203-212.

Nagata, H. (1988). The relativity of linguistic intuition: The effect of repetition on grammaticality judgments. *Journal of Psycholinguistic Research*, 17(1), 1-17.

Onysko, A. (2007). *Anglicisms in German: Borrowing, Lexical Productivity, and Written Codeswitching*. Berlin: Walter de Gruyter.

Pasch, R. (1997). "Weil" mit Hauptsatz—Kuckucksei im "den"-Nest? *Deutsche Sprache*, 25:252-271.

Pfaff, C. (1979). Constraints on language mixing: Intrasentential code-switching and borrowing in Spanish/English. *Language* 55: 291-318.

Pickering, M. J., & H. P. Branigan. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language* 39(4), 633-651.

Poplack, S. (1981). The syntactic structure and social function of code-switching. In R.P. Durán (Ed.), *Latino language and communicative behavior* (pp. 169-184). Norwood, NJ: Ablex.

Rackowsky, A.B. (1989). A study of intra-sentential code-switching in Spanish-English bilinguals and second language learners. Ph.D. dissertation, Brown University.

Reilly, J., Bates, E., & Marchman, V. (1998). Narrative discourse in children with early focal brain injury. *Brain and Language,* 61, 335–375.

Ross, J.R. (1967). Constraints on variables in syntax. Ph.D. dissertation, MIT.

Ross, J.R. (1979). Where's English? In C.J. Fillmore, D. Kemper, & W.S. Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 127-163). New York: Academic Press.

Sankoff, D., &Poplack, S. (1981). A formal grammar for code-switching. *Papers in Linguistics*, 14, 3-45.

Santorini, B., & Mahootian, S. (1995). Codeswitching and the syntactic status of adnominal adjectives. *Lingua* 96: 1-27.

Scheutz, H. (1998). *Weil*-Sätze im gesprochenen Deutsch. In: Hutterer, C. & G. Pauritsch (Eds.). *Beiträge zur Dialektologie des Oberdeutschen Raumes* (pp. 85-112). Göppingen: Kümmerle.

Scheutz, H. (2001). On Causal Clause *Combining*: The Case of weil in Spoken German. In: *Studies in Interactional Linguistics*, edited by Margret Selting & Elizabeth Couper-Kuhlen. Amsterdam: John Benjamins.

Schütze, C.T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago, Il: University of Chicago Press.

Schütze, C.T. (2005). Thinking about what we are asking speakers to do. In Kepser & Rice, 2005, 457-484.

Schütze, C.T. (2009). Web searches should supplement judgments, not supplant them. *Zeitschrift für Sprachwissenschaft*, 28, 151-156.

Sobin, N. (1984). On code-switching inside NP. *Applied Psycholinguistics*, 5(4), 393-303.

Sorace, A. (1988). Linguistic intuitions in interlanguage development: The problem of indeterminacy. In J. Pankhurst, M. Sharwood, & P. Van Buren (Eds.), *Learnability and Second Languages: A Book of Readings* (pp. 167-190). Dordrecht: Foris.

Snow, C., & Meijer, G. (1977). On the secondary nature of syntactic intuitions. In S. Greenbaum (Ed.), *Acceptability in language* (pp. 163-177). The Hague, the Netherlands: Mouton.

Sprouse, Jon. (2007). Continuous Acceptability, Categorical Grammaticality, and Experimental Syntax. *Biolinguistics*, 1: 118-129.

Sternefeld, W. (2000). Grammatikalität und Sprachvermögen. In P. Suchsland, J. Bayer, & C. Römer (Eds.), *Von der Philologie zur Grammatiktheorie: Peter Suchsland zum 65. Geburtstag*. (pp. 15-44). Tübingen: M. Niemeyer.

Stokes, W. (1974). All of the work on quantifier-negation isn't convincing. In M.W. La Galy, R.A. Fox, & A. Bruck (Eds.), *Papers from the tenth regional meeting, Chicago Linguistic Society* (pp. 692-700). Chicago: Chicago Linguistic Society.

Timm, L. (1975). Spanish-English code-switching. El porqué y how-not-to. *Romance Philology*, 28(4), 473-482.

Toribio, A.J. (2001a). Accessing bilingual code-switching competence. *International Journal of Bilingualism*, 5(4), 403-436.

Toribio, A.J. (2001b). On the emergence of bilingual code-switching competence. *Bilingualism: Language and Cognition*, 4(3), 203-231.

Treffers-Daller, J. (2009). Code-switching and transfer: An exploration of similarities and differences. In Bullock & Almeida, 2009, 58-74.

Uhmann, S. (1998). Verstellungsvariationen in *weil*-Sätzen. *Zeitschrift für Sprachwissenschaft*, 17, 92-139.

Valdés, G. (1981). Code switching as deliberate verbal strategy: A microanalysis of direct and indirect requests among bilingual Chicano speakers. In R.P. Duran (ed.) *Latino language and communicative behavior.* Norwood, NJ: Ablex. 95-107.

van Gelderen, E., & MacSwan, J. (2008). Interface conditions and code-switching: Pronouns, lexical DPs, and checking theory. *Lingua*, 118(6), pp. 765-776.

Watzinger-Tharp, J. (2006). German weil-Clauses: Current Research and its Implications for the L2 Classroom. *Die Unterrichtspraxis/Teaching German*, 39(1/2): 46-54. http://web.ebscohost.com.login.ezproxy.library.ualberta.ca/ehost/detail?vid=1&hid=112&sid=07dafdff-8cff-410b-a393-dc931b263d97%40sessionmgr112&bdata=JmxvZ2lucGFnZT1Mb2dpbi5hc3Amc2l0ZT1laG9zdC1saXZlJnNjb3BlPXNpdGU%3d#db=ehh&AN=23901011 (accessed Jan. 31, 2011).

Wei, L. (2000). *The bilingualism reader*. London: Routledge.

Wentz, J. (1977). Some considerations in the development of a syntactic description of code-switching. Ph.D. dissertation, University of Kansas.

APPENDIX A

SURVEY-INFORMED CONSENT GROUP A

Dear Participant,

I am a graduate student under the direction of Professor Dr. MacSwan and Dr. Daniel Gilfillan in the College of Liberal Arts and Sciences at Arizona State University.

I am conducting a research study to determine whether, and to what degree, code-switched statements are acceptable to German-English bilinguals such as yourself. I am writing to ask for your help by participating in a study that will help me develop a new standard in
the experimental protocol for research on bilingualism.

Your participation in this study is voluntary and will involve taking a brief two-part survey (approximately 20 minutes) taken 2 weeks apart. You can skip questions if you wish. If you choose not to participate or to withdraw from the study at any time, there will be no penalty. There are no foreseeable risks or discomforts to your participation.

Participating in the study will not require you to submit your name, and as such, responses will be confidential. The results of this study may be used in reports, presentations, or publications but your name will not be known. In addition, results will only be shown in the aggregate form.

If you have any questions concerning the research study, please contact the research team at: macswan@asu.edu and jgrabow1@asu.edu. If you have any questions about your rights as a subject/participant in this research, or if you feel you have been placed at risk, you can contact the Chair of the Human Subjects Institutional Review Board, through the ASU Office of Research Integrity and Assurance, at (480) 965-6788.

Return of the questionnaire will be considered your consent to participate. The survey may be accessed here:
https://www.surveymonkey.com/s/GSHK5HP

With kind regards,

Jane Grabowski

APPENDIX B

SURVEY-INFROMED CONSENT GROUP B

Dear Participant,

I am a graduate student under the direction of Professor Dr. MacSwan and Dr. Daniel Gilfillan in the College of Liberal Arts and Sciences at Arizona State University.

I am conducting a research study to determine whether, and to what degree, code-switched statements are acceptable to German-English bilinguals such as yourself. I am writing to ask for your help by participating in a study that will help me develop a new standard in the experimental protocol for research on bilingualism.

Your participation in this study is voluntary and will involve taking a brief survey (approximately 20 minutes). You can skip questions if you wish. If you choose not to participate or to withdraw from the study at any time, there will be no penalty. There are no foreseeable risks or discomforts to your participation.

Participating in the study will not require you to submit your name, and as such, responses will be confidential. The results of this study may be used in reports, presentations, or publications but your name will not be known. In addition, results will only be shown in the aggregate form.

If you have any questions concerning the research study, please contact the research team at: macswan@asu.edu and jgrabow1@asu.edu. If you have any questions about your rights as a subject/participant in this research, or if you feel you have been placed at risk, you can contact the Chair of the Human Subjects Institutional Review Board, through the ASU Office of Research Integrity and Assurance, at (480)965-6788.

Return of the questionnaire will be considered your consent to participate. The survey may be accessed here:

https://www.surveymonkey.com/s/RLMHXYQ

With kind regards,

Jane Grabowski

APPENDIX C

SURVEY-DEMOGRAPHIC INFORMATION

This questionnaire is designed to give us a better understanding of your experience learning and using languages. Please be as accurate and thorough as possible when answering the following questions.

1. What was the first language you learned at home?
   a. English
   b. German
   c. Both English and German at the same time
   d. Other (please specify)

2. At what point was literacy in GERMAN initiated?
   a. Elementary school or equivalent
   b. High school or equivalent
   c. University or equivalent

3. At what point was literacy in ENGLISH initiated?
   a. Elementary school or equivalent
   b. High school or equivalent
   c. University or equivalent

4. In which language do you have greater SPEAKING proficiency?
   a. German
   b. English
   c. Same for both English and German

5. In which language do you have greater READING proficiency?
   a. German
   b. English
   c. Same for both English and German

APPENDIX D

PRIMING JUDGMENT TASKS

1. I like that bar because the drinks are cheap.
   a. Completely acceptable and natural
   b. Acceptable, but perhaps somewhat unnatural
   c. Doubtful, but perhaps acceptable
   d. Worse, but not totally unacceptable
   e. Thoroughly unacceptable
   f. Horrible

2. My hands are dirty because I some yard work did.
   a. Completely acceptable and natural
   b. Acceptable, but perhaps somewhat unnatural
   c. Doubtful, but perhaps acceptable
   d. Worse, but not totally unacceptable
   e. Thoroughly unacceptable
   f. Horrible

3. I dropped the plate because full hands my were.
   a. Completely acceptable and natural
   b. Acceptable, but perhaps somewhat unnatural
   c. Doubtful, but perhaps acceptable
   d. Worse, but not totally unacceptable
   e. Thoroughly unacceptable
   f. Horrible

4. Er bleibt zu Hause, weil er eine Erkältung hat.
   a. Completely acceptable and natural
   b. Acceptable, but perhaps somewhat unnatural
   c. Doubtful, but perhaps acceptable
   d. Worse, but not totally unacceptable
   e. Thoroughly unacceptable
   f. Horrible

5. Ich trinke Wasser, weil ich habe Durst.
   a. Completely acceptable and natural
   b. Acceptable, but perhaps somewhat unnatural
   c. Doubtful, but perhaps acceptable
   d. Worse, but not totally unacceptable
   e. Thoroughly unacceptable
   f. Horrible

6. Du sollst zum Arzt gehen, weil Fieber du hast.
   a. Completely acceptable and natural
   b. Acceptable, but perhaps somewhat unnatural
   c. Doubtful, but perhaps acceptable
   d. Worse, but not totally unacceptable
   e. Thoroughly unacceptable
   f. Horrible

7. Ich ziehe eine Jacke an because es heute kalt ist.
   a. Completely acceptable and natural
   b. Acceptable, but perhaps somewhat unnatural
   c. Doubtful, but perhaps acceptable
   d. Worse, but not totally unacceptable
   e. Thoroughly unacceptable
   f. Horrible

8. Ich bin zur Apotheke gegangen because ich wollte Aspirin kaufen.
   a. Completely acceptable and natural
   b. Acceptable, but perhaps somewhat unnatural
   c. Doubtful, but perhaps acceptable
   d. Worse, but not totally unacceptable
   e. Thoroughly unacceptable
   f. Horrible

9. I am waiting for the bus, weil I have don't car a.
   a. Completely acceptable and natural
   b. Acceptable, but perhaps somewhat unnatural
   c. Doubtful, but perhaps acceptable
   d. Worse, but not totally unacceptable
   e. Thoroughly unacceptable
   f. Horrible

APPENDIX E

GRAMMATICALITY JUDGMENT TASKS

1. I'm broke because meine Katze trinkt viel Milch.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

2. I got a good grade because ich habe stundenlang gelernt.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

3. I don't eat meat because ich bin Vegetarier.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

4. I'm broke because meine Katze viel Milch trinkt.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

5. I got a good grade because ich stundenlang gelernt habe.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

6. I don't eat meat because ich Vegetarier bin.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

7. Ich habe kein Geld because meine Katze trinkt viel Milch.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

8. Ich habe eine gute Note bekommen, because ich habe studenlang gelernt.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

9. Ich esse kein Fleisch because ich bin Vegetarier.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

10. Ich habe kein Geld because meine Katze viel Milch trinkt.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

11. Ich habe eine gute Note bekommen because ich stundenlang gelernt habe.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

12. Ich esse kein Fleisch because ich Vegetarier bin.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

13. I'm broke, weil my cat drinks a lot of milk.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

14. I got a good grade, weil I studied for hours.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

15. I don't eat meat, weil I am a vegetarian.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

16. I'm broke, weil my cat a lot of milk drinks.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

17. I got a good grade, weil I for hours studied.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

18. I don't eat meat, weil I a vegetarian am.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

19. Ich habe kein Geld, weil my cat drinks a lot of milk.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

20. Ich habe eine gute Note bekommen, weil I studied for hours.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

21. Ich esse kein Fleisch, weil I am a vegetarian.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

22. Ich habe kein Geld, weil my cat a lot of milk drinks.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

23. Ich habe eine gute Note bekommen, weil I for hours studied.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

24. Ich esse kein Fleisch, weil I a vegetarian am.
    a. Completely acceptable and natural
    b. Acceptable, but perhaps somewhat unnatural
    c. Doubtful, but perhaps acceptable
    d. Worse, but not totally unacceptable
    e. Thoroughly unacceptable
    f. Horrible

APPENDIX F

IRB APPROVAL

Office of Research Integrity and Assurance

To:	Elly Van Gelderen
	LL
From:	Mark Roosa, Chair
	Soc Beh IRB
Date:	02/25/2011
Committee Action:	**Exemption Granted**
IRB Action Date:	02/25/2011
IRB Protocol #:	1102006070
Study Title:	On the Stability of Grammaticality Judgments in German-English Code-Switching

The above-referenced protocol is considered exempt after review by the Institutional Review Board pursuant to Federal regulations, 45 CFR Part 46.101(b)(2) .

This part of the federal regulations requires that the information be recorded by investigators in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects. It is necessary that the information obtained not be such that if disclosed outside the research, it could reasonably place the subjects at risk of criminal or civil liability, or be damaging to the subjects' financial standing, employability, or reputation.

You should retain a copy of this letter for your records.