2020-06-17

A Corpus-Based Study of the Gender Assignment of Nominal Anglicisms in Brazilian Portuguese

Taryn Marie Skahill
Brigham Young University

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Arts and Humanities Commons

A Corpus-Based Study of the Gender Assignment of

Nominal Anglicisms in Brazilian Portuguese

Taryn Marie Skahill

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Arts

Earl K. Brown, Chair
Cynthia Hallen
Mark Davies
Dan Dewey

Department of Linguistics

Brigham Young University

ABSTRACT

A Corpus-Based Study of the Gender Assignment of
Nominal Anglicisms in Brazilian Portuguese

Taryn Marie Skahill
Department of Linguistics, BYU
Master of Arts

The purpose of this study is to analyze the variability of gender assignment to nominal anglicisms in Brazilian Portuguese and to identify how the orthography of English loanwords and their establishment in the language influences such variation. This study also seeks to identify the most important factors that govern such gender assignment. The data were gathered from two Portuguese corpora, one consisting of more formal and edited language (Corpus do Português, News on the Web) and the other consisting of less formal and unaltered language, such as blog posts (Corpus do Português, Web/Dialects). Forty anglicisms were analyzed in order to study the variation in gender assignment based on the anglicisms' orthography and establishment in the language, as well as to help determine whether the gender of the loanwords' cognates or calques influences the gender assignment of words borrowed from English into Portuguese. The results of this study indicate that the gender of the anglicisms' cognates or most frequent calques and the gender found in Portuguese dictionaries equally influence the gender assignment to anglicisms. This research also shows that variability in gender assignment is not significantly affected by an English loanword's attestation in Portuguese dictionaries nor by the adaptation to Portuguese orthography of English loanwords, though there are trends that indicate a negative correlation between the variability of anglicisms' genders and their attestation in Portuguese dictionaries.

ACKNOWDGEMENTS

Many people have influenced my academic career and are to thank for their contributions, whether direct or indirect, to the completion of this thesis. I express gratitude to the faculty members at BYU who helped me explore linguistics and learn more about this field that I find fascinating. These faculty members include, but are not limited to: Dr. Bill Eggington, Dr. Wendy Smemoe, Dr. Chris Rogers, Dr. Norm Evans, Dr. Grant Eckstein, and Mary Beth Wald.

Professors who I did not include in my previous list, but who have undoubtedly made a profound impact on my graduate studies, and particularly my Master's thesis, are Dr. Dan Dewey, Dr. Mark Davies, Dr. Cynthia Hallen, and my chair, Dr. Earl Brown. I want to thank Dr. Dan Dewey for influencing me with his humble disposition and passion for language acquisition. I want to thank Dr. Mark Davies for developing his plethora of corpora and for teaching me along with my cohort about their many functions. I want to thank Dr. Cynthia Hallen for mentoring me in the field of teaching through her example and for her support when I was deciding on a research topic. Finally, I would like to thank my chair, Earl Brown, for the innumerable hours he spent coaching me in person and through Zoom as I continuously sought to refine my research. His patience, discipline, and encouragement went a long way in assisting me with the completion of this project.

TABLE OF CONTENTS

1 INTRODUCTION

Language contact is one of several factors that causes languages to constantly change (other factors that drive change include time, language planning, etc.). This contact can occur face-to-face and in person as language groups reside amongst each other, but technology has greatly facilitated the spread of language and the influence one language has on another. Technology makes language contact easier to accomplish and more widespread. One effect of language contact is borrowing, while other effects include creolization or language extinction.

In borrowing situations, the language into which a word is borrowed often must account for certain grammatical features that differ from the language from which a word was borrowed. One example is when Spanish and Portuguese borrow verbs from English—the speakers of these two languages often may "-ar" or "-er" to the end of verbs borrowed from English to make them more similar to native Spanish and Portuguese verbs (e.g., the English slang term 'to ship' means to support a relationship, and in Portuguese this verb has become *shipar*), and to give them a form that can be conjugated (e.g., *shipo* in Portuguese means 'I ship' or 'I support that relationship'). Likewise, when languages that have grammatical gender borrow words from languages that do not use grammatical gender, a grammatical gender must be assigned to the borrowed words. That is what this study analyzes with regard to English words receiving a grammatical gender in Portuguese. These recipient languages must adapt the grammar structures and apply them to the borrowed language features in order to accommodate the words and phrases incorporated from other languages. The way that native speakers, whether consciously, or more likely subconsciously, decide to do this may vary individually, thus resulting in variation of how loanwords are depicted in the new language. In other words, due to native speakers' lack of familiarity with a new word borrowed from another language, they may adapt it to the new

language's features in different ways, based on different rules or personal preferences or experiences.

When words are borrowed from a language that does not use nominal grammatical gender into a language that does assign a grammatical gender to nouns and adjectives, the speakers of the recipient language must distinguish which gender the borrowed words should receive. Different languages have different gendering systems (e.g., masculine and feminine in Portuguese and Spanish, neuter and common in Dutch, etc.). Due to speakers' individual differences and experiences with their native language, this gender assignment could vary, resulting in a lack of consistency in the genders that are assigned to loanwords and leading to lack of understanding of which gender should actually be assigned to words borrowed from other languages. What this study seeks to identify are the factors that influence variability amongst the genders assigned to loanwords borrowed from English into Brazilian Portuguese, as well as the factors that influence gender assignment.

The purpose of this thesis is to analyze the variation in and identify the factors that govern gender assignment to nominal anglicisms in Brazilian Portuguese. The two languages in this study are English (the donor language) and Brazilian Portuguese (the recipient language). Thus, the loanwords in this study may be referred to as anglicisms. Research has been done in regards to how and why anglicisms are used in Brazilian advertising and brand naming (Friedrich, 2002), Brazilian websites (Diniz de Figueiredo, 2010), and more specifically in Brazilian fashion blogs (Ribeiro dos Santos, 2016). These articles are informational as to the usage of anglicisms in Brazilian media, but give no insight regarding the reasoning behind the genders affiliated with these loanwords, nor the loanwords' levels of variability in gender assignment. Research concerning variation in gender assignment has been done for Dutch

(Franco, et. al, 2018), Italian (Rabeno & Repetti, 1997), Spanish (Clegg & Waltermire, 2009), German (Onysko, et. al, 2013; Gilbert, G. G.,1965), and other languages with grammatical gender in various settings throughout the world. These studies have identified a wide variety of factors that influence which genders are most commonly assigned to anglicisms. Such factors include the gender of semantic categories (Geerts, 1996 as cited by Franco, et. al, 2018), the gender of synonyms in the host language (Geerts, 1996 as cited by Franco, et. al, 2018 as cited by Franco, et. al, 2018; Clegg & Waltermire, 2009), genders associated with suffixes (Geerts, 1996 as cited by Franco, et. al, 2018) or terminal phonemes (Clegg & Waltermire, 2009), and default genders assigned to essentially all loanwords (Clegg & Waltermire, 2009). These studies were conducted specifically for other languages (e.g., Dutch, Spanish, etc.) and how the speakers of these languages assign gender to words that have been borrowed from English. However, there are no available studies on the gender assignment of nominal anglicisms in Brazilian Portuguese. Thus, it is uncertain as to whether these previously mentioned factors influence the gender assignment to nominal anglicisms in Brazilian Portuguese the same way that they influence other languages, or if there exists an entirely different set of variables that influence gender assignment to English loanwords in Portuguese. Which specific factors most strongly lead to variability in gender assignment to anglicisms used in Portuguese are also currently unknown. Therefore, this study helps discover the factors that affect which gender is most frequently assigned, but also what factors influence variability amongst gender assignment. Using data collected from two Portuguese corpora and information from four Portuguese dictionaries, this study seeks to determine what factors lead to higher variability in masculine or feminine gender assignment to nominal anglicisms in Brazilian Portuguese.

2 REVIEW OF LITERATURE

In this discussion we will explore the past research already conducted across different languages on variability in gender assignment to anglicisms. We will first look at borrowing in a general sense. Then, we will identify some potential reasons that borrowing across languages occurs. Finally, we will present the findings from key articles after which this study was modeled. These studies highlight the variables that influence gender assignment in Dutch and Spanish.

2.1 Borrowing Defined

Myers-Scotton (2002) quotes Haugen's (1950) definition of borrowing as, "the attempted reproduction in one language of patterns previously found in another." Myers-Scotton (2002) clarifies by saying, "By 'pattern' Haugen means largely lexical elements, not grammatical patterns" (p. 234). The English loanwords, or anglicisms, analyzed in this study are lexical elements. Myers-Scotton also discusses the research of Thomason and Kaufman (1988) on borrowing. She shares their definition of borrowing as well, which is "the incorporation of foreign features into a group's native language by speakers of that language: the native language is maintained but is changed by the addition of the incorporated features." The incorporated feature in this study is gender assignment. Thomason and Kaufman (1988), as cited by Myers-Scotton (2002), also say that "Borrowing is differentiated from interference in that speakers who borrow elements still retain their L1." Thus, the anglicisms to be analyzed in this study are English words used amidst Portuguese phrases, and nominal gender has been added to these anglicisms.

Myers-Scotton (2002) explains that "the most obvious statistic about borrowing is that more nouns are borrowed than any other category" (p. 240). After citing other studies whose authors indicate a higher number of loanwords that were nouns as opposed to other parts of speech in the languages they were researching, she goes on to say, "This statistic is no surprise because almost all borrowed elements are content morphemes, and nouns, along with verbs, are the prototypical content morphemes" (p. 240). She hypothesizes that:

> "Nouns are borrowed more frequently than any other category because they receive, not assign, thematic roles. That is, their insertion into the frame of another language is less disruptive of predicate-argument structure than insertions of any content morphemes assigning thematic roles (i.e., verbs, but also prepositions and predicate adjectives)" (Myers-Scotton, 2002).

Consequently, all anglicisms studied in this thesis are nouns, because only nouns receive a gender. Adjectives and determiners are also gendered, but their gender depends on the noun they modify. In this study and others, genders of the adjectives and determiners are how the gender of the loanword nouns is identified.

Haspelmath and Tadmor (2009) cite Myers-Scotton (2002), who presents a dichotomy between types of loanwords in which she includes "cultural borrowings" and "core borrowings" (p. 46). Cultural borrowings are defined as "words for objects and concepts new to the culture" as in "hard drive, SUV, and global warming" (Myers-Scotton, 2002, p. 41). These are words that do not necessarily have a translation equivalent, so the idea or object itself may be new to the host culture. These words are borrowed because the recipient language does not have a way to express it. Because the concept is new to the speakers of the recipient language, these words do not have a synonym in the recipient language. Differently, core borrowings are defined as "words that more or less duplicate already existing words in the L1" as in *le weekend* 'the weekend' adapted into French (as opposed to *fin de semaine*) (Myers-Scotton, 2002, p. 41).

These words do have a translation equivalent in the recipient language, but the speakers of the language still prefer to borrow another language's iteration for it.

Friedrich (2002) suggests that English words are borrowed into other languages for a variety of reasons. Some of these reasons include: its feeling of prestige (e.g., a cleaning-service contractor in Brazil called 'Master Limp,'; "Limp" is derived from the Portuguese word *limpar*, meaning 'to clean'); the target culture's desire for Westernization (e.g., referring to clothing styles as *fashion*, or using the American phrase *fake news*); a preference for shorter words in some cases (e.g., referring to someone's *namorado* or 'boyfriend' as *boy*); the need for precision in the naming of new objects (which would be considered a cultural borrowing) (e.g., *calça jeans* 'jean pants'); or just a desire for innovation and creativity (e.g., referring to a billboard as *outdoor*). Though languages borrow words for many reasons, the purpose of this thesis is not to analyze the motive behind borrowing from language to language, but rather to highlight the factors that influence gender assignment to anglicisms in Brazilian Portuguese, and what creates variation in this gender assignment.

2.2 Definition of Terms

In this study, the words "variability" and "variation" are used interchangeably, with no discrepancy in meaning. "Anglicism" refers to a loanword borrowed from the English language. The word 'cognate' is used to describe a word in another language that is similar in spelling and pronunciation, and is identical in meaning across the two languages. For example, the Portuguese word *mensageiro* is the cognate of the English word 'messenger' because they mean the same thing and they are similar in spelling and pronunciation. The word 'calque' is used to mean the translation equivalent. If an anglicism is a core borrowing, it is likely that there is a synonym in Portuguese, and that is considered the calque. For example, *fake news* can be directly translated

to mean *notícias falsas*, and that would be considered its calque. If an anglicism is a cultural borrowing, then there is likely no direct translation, and a short definition of the word in Portuguese is considered the calque. For example, the word 'iPhone' in Portuguese could be calqued as *telefone celular da marca de Apple* ('cellular phone from the brand Apple).

The Portuguese spellings of words are written in italics throughout this thesis. Their English glosses are written in single quotes. When applicable, the part of speech of the Portuguese word follow in small caps. When a loanword is described in an English setting, it can be found in single quotes. When a the loanword is described for a Portuguese setting, it is italicized.

2.3 Previous Research

Research has been conducted in regards to gender assignment to anglicisms in other languages, including Dutch, German, Italian, and Spanish (Franco, et. al, 2018; Onysko, et. al, 2013; Gilbert,1965; Rabeno & Repetti, 1997; Clegg & Waltermire, 2009). Research has also been done regarding other aspects of anglicisms in Brazilian Portuguese (i.e., usage of English loanwords in advertising, brand naming, fashion blogs, etc.) (Friedrich, 2002; Diniz de Figueiredo, 2010; Ribeiro dos Santos, 2016). Much research has also been conducted on loanwords in general. Franco, et. al (2018) as well as Clegg and Waltermire (2009) conducted research on the gender assignment of anglicisms in Dutch and Spanish, respectively. Both studies aim to discover the underlying factors that govern the way English loanwords receive a gender when they are borrowed into a new language.

The study done by Franco, et. al (2018) on English loanwords in Dutch found that "Anglicisms are assigned common gender by default unless there is a reason for using neuter," which reasons "can be semantic or morphological, or they can involve some kind of analogy

7

between the anglicism and a native Dutch noun." They explain that "the gender of nouns in Dutch is largely arbitrary" and that the two options for gender assignment in Dutch are common and neuter (Franco, et. al, 2018). The authors cite Geerts (1996), who presents a hierarchy of factors that influence gender assignment to loanwords. He claims that "the gender of a cognate has a stronger influence than the prototypical gender of nouns within the same semantic field; the latter, in turn has a stronger influence than the gender associated with a particular suffix, which is, again, more important than a translation equivalent's gender" (Franco, et. al, 2018). For example, Franco, et. al (2018) explain that "*gingerbeer* is a neuter noun in Dutch because its association with the native neuter cognate *bier* 'beer' takes precedence over the prototypical common gender associated with the semantic field 'drinks.'" Since the loanword has a cognate, its gender assignment is based on that rather than on the semantic category in which it falls, because cognate is higher on the hierarchy than semantic field is.

Thus, the hierarchy presented by Geerts is as such:

1. Cognate

2. Same semantic field

3. Gender associated with a particular suffix

4. Translation equivalent

This hierarchy "describes what happens if multiple rules apply to a single anglicism in Dutch" (Franco, et. al, 2018). Essentially what Geerts is implying is that if the anglicism has a cognate, the anglicism adopts the cognate's gender. If there is no cognate, the anglicism adopts the gender that is typical of the semantic category. If there is no typical semantic grouping, then the anglicism adopts the gender associated with a particular gender-denoting suffix. And finally, if the anglicism has no such suffix, then it takes the gender of the translation equivalent, or

8

calque. None of the anglicisms in this study belong to a semantic field or possess a suffix that Portuguese grammar rules typically associate with a particular gender, so these factors as presented in Geerts' hierarchy are not relevant to this particular dataset.

Clegg and Waltermire (2009) seek to analyze the gender assignment of English loanwords in Spanish spoken in the southwestern United States. They explain that "the two prevailing views" in regards to gender assignment of anglicisms "are that the process is either done by default where masculine gender is simply assigned to all nouns or a natural process where gender is assigned using the same criteria as that of native Spanish nouns" (p. 1). Their study seeks to determine whether this applies to their data and which factors influence gender assignment to anglicisms in Spanish. Poplack (1993), as quoted by Clegg & Waltermire, defines borrowing as "the adaptation of lexical material to the morphological and syntactic (and usually phonological) patterns of the recipient language" (p. 2). The loanwords in this study are lexical, and the morphological and syntactic pattern to which they are being adapted is grammatical gender.

Clegg and Waltermire (2009) also describe how the grammar of Spanish can help determine whether the anglicism is considered feminine or masculine. They explain that "single English-origin nouns in otherwise Spanish discourse are frequently preceded by a Spanish article" (p. 3). The same is true for Portuguese, and such articles indicate the gender of the word; in Portuguese, the article's gender must match the noun's gender. Definite articles (i.e., *o* 'the' DEF ART MASC and *a* 'the' DEF ART FEM) and indefinite articles (i.e., *um* 'a' 'an' INDEF ART MASC and *uma* 'a' 'an' INDEF ART FEM) can help distinguish masculine and feminine anglicisms. Adjectives as well, both possessive and descriptive, can help determine an English loanword's gender in Portuguese. Possessive masculine adjectives include *meu* 'my' POSS PRO MASC, *seu* 'your' POSS

PRO MASC, *teu* 'your' POSS PRO MASC , and *nosso* 'our' POSS PRO MASC. Possessive feminine adjectives include *minha* 'my' POSS PRO FEM, *sua* 'your' POSS PRO FEM, *tua* 'your' POSS PRO FEM, and *nossa* 'our' POSS PRO FEM. These aforementioned possessive adjectives form a closed class, so it is possible to give an exhaustive list. However, the following descriptive adjectives are in an open class, so an exhaustive list is unrealistic. Some, some examples of descriptive masculine adjectives include *lindo* 'handsome' ADJ MASC, *barato* 'cheap' ADJ MASC, *duro* 'hard' or 'stiff' ADJ MASC, and *novo* 'new' ADJ MASC. Some descriptive feminine adjectives include *linda* 'pretty' ADJ FEM, *ótima* 'great' ADJ FEM, *clara* 'clear' ADJ FEM, and *limpa* 'clean' ADJ FEM.

Using the research of Poplack and Sankoff (1982), Clegg and Waltermire (2009) identify the factors that have the strongest influence on gender assignment to anglicisms in Spanish spoken in the southwest United States. The first factor they discuss, and that which they claim to be most relevant, is that of biological gender (p. 5). This factor, however, is irrelevant in my study because anglicisms that do not have biological gender were intentionally chosen. Some English loanwords used in Brazilian Portuguese (e.g., *coach*, *boy*, *dog*, *friend*, etc.) do have a biological gender. These were decidedly eliminated from the study because it is obvious that the biological gender of the person the anglicism describes is what indicates which grammatical gender is used. For example, if someone in Portuguese said, *ela é minha friend*, 'she is my friend,' it is clear that the anglicism *friend* in this case is feminine because it refers to a female, so the assigned grammatical gender is feminine. If a Portuguese speaker said, *gosto do meu coach*, 'I like my coach,' the person's coach is obviously a man.

Clegg and Waltermire (2009) explain that if an English loanword lacks a biological gender, then the anglicism's terminal phoneme determines the gender (p. 6), which is similar to Geerts' (1996, as cited by Franco, et. al, 2018) criterion of suffix. In Spanish (as studied by

Clegg and Waltermire) as well as its fellow Romance language Portuguese, the terminal phoneme "of a word is an excellent predictor of grammatical gender" (2009). However, there are many exceptions, and very few of the loanwords included in this study end in phonemes that are generally determinant of a word's gender in Portuguese, such as '-o' for masculine and '-a' for feminine. Importantly, Clegg and Waltermire deduced that "synonyms do not play a crucial role in gender assignment." They claim this because in their study, "there were... numerous tokens that had no clear synonym at all. If synonymic gender were a vital part of gender assignment, it would be difficult to assign a gender to words without a synonym." They also point out, "Occasionally, nouns lack... an obvious synonym... Also, several single nouns of English origin... have two synonyms in Spanish, one masculine and one feminine..." (p. 6). For example, "there are two possible Spanish synonyms for the word 'town', *la ciudad* or *el pueblo*" (Clegg and Waltermire, 2009). These two calques in Spanish have different genders, so there would be no way to assign one specific gender to the anglicism 'town.' Clegg and Waltermire (2009) explain that, "Nouns that do not end in these typical phonemes and have no biological gender will be assigned masculine gender demonstrating the preference for Spanish phonological rules in gender assignment" (p. 14). So while Geerts (1996, as cited by Franco, et. al, 2018) claims that when an anglicism's suffix does not indicate gender assignment, the translation equivalent plays a role, Clegg and Waltermire (2009) claim that the gender of the translation synonym does not have an impact but rather that words in Spanish are assigned masculine gender by default. Finally, English loanwords that are "cultural borrowings," or loanwords that embody a new concept in the language as defined by Myers-Scotton (2002), may not have a synonym or equivalent in the target language. This is in reference to the many loanwords that are

borrowed simply because the host language does not really have another way to express the concept.

The findings of Geerts (1996, as cited by Franco, et. al, 2018) and Clegg and Waltermire (2009) seem to contradict each other in regards to the significance of translation equivalent, or calque. Geerts found that translation equivalent does influence gender assignment to anglicisms, but only when no other factors are present. On the other hand, Clegg and Waltermire (2009) share Poplack and Sankoff's (1982) view that translation equivalent does not "play a crucial role in gender assignment." Of course, this is in regards to the respective languages that they studied; Dutch has a different gendering system than Spanish does. Dutch grammar has common and neuter gender, while Spanish grammar has masculine and feminine gender (as does its Romance counterpart, Portuguese). One purpose of this study is to discover whether the gender of an anglicism's calque does have a significant influence on the gender most commonly assigned to the anglicism.

2.4 Research Questions and Hypothesis

The research questions for this study are as follows:

Table 1: Research Questions

1. What effect does an anglicism's level of attestation in Portuguese dictionaries have on the consistency of gender assignment?
2. Do anglicisms that have changed orthographically towards Portuguese orthography show less variation in gender assignment than anglicisms that have not undergone an orthographic change?
3. Does the gender used for anglicisms by the majority of people as represented in corpus data agree with the gender of the cognate or most frequent calque?
4. For the anglicisms that appear in at least one Portuguese dictionary, do the findings from corpus data accurately depict what the dictionaries indicate that the gender should be?

2.4.1 Hypotheses

For research question 1, it is hypothesized that there is a negative correlation—the more attested an anglicism is in Portuguese dictionaries (the more dictionaries in which an anglicism appears), the less variation there would be in gender assignment. This was hypothesized because when an anglicism becomes more established in the language, it seems that the grammatical gender would be more unified amongst speakers, and attestation in the dictionary is a good indicator of an anglicism's establishment in the language.

For research question 2, it is hypothesized that more variation in the gender assignment from the corpus data exists among the original English spellings and that words that have been altered orthographically would show less variation in gender assignment. This was hypothesized because it seems like Portuguese speakers would assign gender more uniformly to words that are spelled with the orthography that matches their native pronunciation.

For research question 3, it is hypothesized that the gender of the cognate or calque has a significant effect on the gender of the anglicism. This is because in Clegg and Waltermire's study, they agree with Poplack and Sankoff (1982) that first biological gender, then terminal phoneme are what most strongly influence gender assignment. If neither of those factors are present in an anglicism, then the default masculine gender is assigned. Their study shows that synonymic gender does not affect the gender assigned to English loanwords. However, Spanish phonology is more similar to English phonology than Portuguese phonology is to English phonology. Therefore, it is possible that the terminal phonemes of Spanish words that indicate gender appear more at the end of English loanwords than the terminal phonemes of Portuguese do. For example, many of the suffixes in Portuguese that indicate a particular gender are nasal vowel sounds, which are not used in the English phonetic system. Clegg and Waltermire (2009)

say that "Spanish is comprised of a greater number of frequent high productivity terminal phonemes, /*n*, *o*, *r*, *s*, *e*, *l*/" for masculine suffixes, and there are a greater number of English words that end in these phonemes than there are English words that end in nasal vowels or any of the suffixes that are typically associated with a gender in Portuguese (see Table 2 in section 3.2). Therefore, it seems that terminal phonemes in Spanish are more likely to affect the gender assignment of English loanwords because these anglicisms are more likely to have terminal phonemes that are more similar to Spanish than to Portuguese. And because so many terminal phonemes of these words are influencing the gender assignment, another option like synonymic gender is not needed. However, because Portuguese does not have masculine suffixes that are similar to any English terminal phonemes, a different factor must be used to indicate gender assignment, and it was hypothesized that this factor would be the cognate or calque gender.

For research question 4, it is hypothesized that data collected from corpora reflects the appropriate genders as indicated by the dictionaries. In other words, it was believed that people would do what the dictionaries say they should be doing in regards to gender assignment of the loanwords in the dataset. This was hypothesized because it seems like an anglicism would become attested in a dictionary after its usage in the language, not the other way around. In other words, it did not seem likely that multiple dictionaries would indicate that a particular gender should be assigned to a certain loanword without there being a great deal of evidence that this gender is what is generally used and accepted.

3 METHODOLOGY

3.1 Anglicism Selection

The decision about which anglicisms to analyze in this study was made by consulting a variety of Brazilian websites and social media posts. An Internet search was conducted for *palavras inglesas que são usadas em portuguê*s ('English words that are used in Portuguese') and returned several lists[1] that were created by native Portuguese speakers who are fluent in—or at least familiar with—the English language. These multiple lists were compiled into one single list and others that had been previously heard or seen were added. At the end of this process, the list ended up with around 128 anglicisms in total. Anglicisms that were not nouns were then eliminated (because in order to receive a gender, the word must be a noun), as well as loanwords that were animate objects (because their gender would obviously be assigned as masculine or feminine based on biological sex.) More loanwords were excluded at random until the list ended up with a total of 40 anglicisms, ten of which include a Portuguese and an English spelling, totaling 30 individual words (10 with two different spellings). Anglicisms from the original list of 128 had to be excluded simply to make this project manageable. Because 50 tokens were scraped for each type, there had to be a reasonable limit, and 40 types with 50 tokens each seemed like a manageable quantity that would also give meaningful results. For this same reason,

---

[1] https://ingleswinner.com/blog/100-palavras-em-ingles-que-usamos-dia-dia/ https://www.mundovestibular.com.br/estudos/ingles/300-palavras-em-ingles-que-conhecemos-por-causa-da-globalizacao https://duvidas.dicio.com.br/palavras-em-ingles-que-usamos-no-dia-a-dia/ https://www.normaculta.com.br/anglicismos-palavras-de-origem-inglesa-na-lingua-portuguesa/
https://www.wizard.com.br/idiomas/entenda-o-que-e-anglicismo-e-confira-80-exemplos/ https://www.culturainglesacuritiba.com.br/anglicismo-do-ingles-direto-para-o-portugues/

only the singular versions of the words were searched and not the plural forms. The complete list of anglicisms is: *bike*, *black*, *blazer*, *blêiser*, *club*, *clube*, *DVD*, *fake news*, *feedback*, *freezer*, *film*, *filme*, *goal*, *gol*, *hobby*, *Internet*, *iPhone*, *kit*, *LAN House*, *like*, *look*, *Messenger*, *notebook*, *outdoor*, *picnic*, *piquenique*, *pet shop*, *print*, *pudding*, *pudim*, *sandwich*, *sanduíche*, *selfie*, *telephone*, *telefone*, *ticket*, *uísque*, *video*, *vídeo*, *whiskey*. A collection of meanings of each of these words as they are used in Portuguese can be found in Appendix 7.3. The two types of borrowings as depicted by Myers-Scotton (2002), both cultural and core, are included in this study of English loanwords in Portuguese because the type of borrowing is not one of the factors analyzed as to how it influences gender assignment.

3.2 Portuguese Grammar

Every Portuguese noun has grammatical gender. Some word-endings pattern with masculine nouns, while some word-endings pattern with feminine nouns. For example, most nouns with word final '-o' have masculine gender, while the majority of nouns with word-final '-a' have feminine gender. However, with some words, obviously including anglicisms, the gender is unclear. Some Portuguese words' gender is not made clear by the word-ending, or are exceptions to the general rules. For example, *dia* 'day' NOUN MASC is a common masculine word ending in the typical feminine suffix, '-a.' Words with Greek etymons (e.g., *problema* 'problem,' *programa* 'program,' *tema* 'theme,' etc.) tend to end in the suffix '-ma' but are masculine (Lourdes Sá Pereira, 1948). *Tribo* 'tribe' NOUN FEM is a feminine word ending in the usually masculine '-o.' Other suffixes such as '-im,' '-om,' '-um,' '-em,' and '-ã' (Perini, 2002) are generally masculine, while the suffixes '-gem,' '-ice,' '-dade,' and '-ção' are generally feminine (Lourdes Sá Pereira, 1948). Table 2 below shows the suffixes that typically indicate feminine and masculine grammatical gender according to Lourdes Sá Pereira (1948).

Table 2: Typical Suffixes that Indicate a Particular Gender in Portuguese

Typical feminine suffixes:	Typical masculine suffixes:
-a	*-o*
-gem	*-im*
-ice	*-om*
-dade	*-um*
-ção	*-em*
	-ã

In words with less common endings such as *parede* 'wall' NOUN FEM and *lugar* 'place' NOUN MASC, the gender is less obvious. Non-native Portuguese speakers must pay attention to the words that precede or follow nominals in which the gender was not based on the word-final sounds. The same principle applies to the determination of gender of anglicisms in Brazilian Portuguese; common prepositional and adjectival collocates as well as the articles used around the anglicisms in the corpus are used to see which gender is most typically assigned to the specific loanword, and how those genders vary.

Certain semantic categories in Portuguese tend to assign different grammatical genders. Names of seas, rivers, lakes, winds, and months are typically masculine (e.g., *o Atlântico* 'the Atlantic' NOUN MASC, *o Amazonas* 'the Amazon' NOUN MASC, *o Mediterrâneo* 'the Mediterranean' NOUN MASC, *o fevereiro* 'February' NOUN MASC, *o furacão* 'the hurricane' NOUN MASC, etc.) (Lourdes Sá Pereira, 1948; Perini, 2002). The names of mountains also generally adopt masculine gender (e.g., *o Himalaia* 'the Himalayas' NOUN MASC, etc.) (Perini, 2002). Names of the letters of the alphabet, numbers, and musical notes are also masculine (*o "b"* 'the "b"' NOUN MASC, *o quatro* 'the four' NOUN MASC, etc.) (Lourdes Sá Pereira, 1948). Names of continents and

islands tend to be feminine (e.g., *Europa* 'Europe' NOUN FEM, *a Sicília* 'Sicily' NOUN FEM) (Lourdes Sá Pereira, 1948). The names of cities also take a feminine gender (e.g., *a Paris* 'Paris') (Perini, 2002). There are exceptions to these semantic categories, but Portuguese nominals have a tendency to receive a specific gender based on these semantic categories.

The masculine gender in Portuguese is represented in a number of ways other than suffixes or semantic groups. It can be inferred that a noun preceded by the articles *o* 'the' DEF ART MASC or *um* 'a' 'an' INDEF ART MASC is masculine. A noun followed by (or preceded in some cases) an adjective ending in the suffix '-o' (e.g., *lindo* 'handsome' ADJ MASC, *barato* 'cheap' ADJ MASC, *duro* 'hard' ADJ MASC, *novo* 'new' ADJ MASC, etc.) is masculine. Possessive pronouns are also gender-specific, so nominals preceded by *meu* 'my' POSS PRO MASC, *seu* 'your' POSS PRO MASC, or *nosso* 'our' POSS PRO MASC are also considered masculine. *Teu* 'your' POSS PRO MASC is another variation of *seu* 'your' POSS PRO MASC, and while it does occur most commonly in Portugal, it is also sometimes used in Brazil. All of these pronouns can be pluralized as well in order to represent the number of objects being possessed. Not all adjectives are gendered (e.g., *legal* 'cool'ADJ, *difícil* 'difficult' ADJ, *feliz* 'happy' ADJ, etc.) so not every adjective is necessarily an indicator of the anglicism's assigned gender.

Nouns preceded by the articles *a* 'the' DEF ART FEM or *uma* 'a' 'an' INDEF ART FEM are feminine. A word that is followed by (or preceded in some cases) an adjective ending in the suffix '-a' are feminine (e.g., *linda* 'pretty' ADJ FEM, *ótima* 'great' ADJ FEM, *clara* 'clear' ADJ FEM, *limpa* 'clean' ADJ FEM, etc.) are feminine. Furthermore, the feminine possessive pronouns *minha* 'my' POSS PRO FEM, *sua* 'your' POSS PRO FEM, or *nossa* 'our' POSS PRO FEM also represent the feminine gender. *Tua* 'your' POSS PRO FEM is another variation of *teu* 'your' POSS PRO FEM. Again, while it does occur most commonly in Portugal, it is also sometimes used in Brazil and identifies

a loanword as having feminine gender. And once again, all of these pronouns and adjectives can be made plural when the noun following them is plural. These collocates help in determining the genders that are most commonly assigned to anglicisms as they are searched in the corpus. A complete list of potential collocates that are used to identify which gender the anglicism has been assigned can be found in Appendix 7.2.

3.3 Factors of Gender Assignment and Variation

This study looks at how anglicism establishment and orthographic adaptation influence variation in gender assignment. It also seeks to identify whether translation equivalent (calque) plays a significant role in the chosen gender assigned to English loanwords. It does not speculate why speakers of Portuguese choose to use anglicisms instead of their Portuguese equivalents, but rather the factors that determine whether masculine or feminine gender is assigned and the reasons for variation in that assignment. In order to determine how established an anglicism is in Portuguese, each English word was looked up in four online Portuguese dictionaries.[2] Words that appear in three or all four of the dictionaries could be considered more established, while words that appeared in one or none of the dictionaries are less established. A complete table of the anglicisms and the dictionaries in which they appear can be found in Appendix 7.4. Another factor that could influence gender assignment is orthographic change. Ten of the words in this study were looked at both with English and Portuguese orthography, and the data were analyzed to discover the spelling's effect on gender assignment.

---

[2] https://dicionario.priberam.org/, https://michaelis.uol.com.br/, https://www.infopedia.pt/, https://www.dicio.com.br

3.4 Corpus Data Extraction Process

Two different corpora were used to collect data: The News on the Web (NOW) corpus from the Corpus do Português and the Web/Dialects corpus from the Corpus do Português (Davies, 2016), both created by Mark Davies[3]. The NOW corpus has edited language, as it is published news. However, the Web/Dialects corpus includes blogs and less formal web posts that have not been edited. Therefore, the data include texts that have been edited as well as texts that are uploaded by the random public, resulting in a balanced distribution of language formality. It is important to note that the Web/Dialects corpus was created in 2014 and has not been updated since its initial version, so any words that have more recently entered the language are not included in the corpus. These two corpora were chosen not only for having plenty of language samples to analyze, but also because they have a user-friendly interface and are free to the public.

To analyze the gender assignment of each word, 25 random samples were collected from each of the two corpora (Corpus do Português, NOW; Corpus do Português, Web/Dialects), specified for Brazil only (as these corpora have data from four Portuguese speaking countries—Mozambique, Angola, Portugal, and Brazil), resulting in 50 random tokens for each of the 40 anglicisms (30 different words, ten of which consisted of two different spellings).[4] Though this is a relatively low number of tokens compared to the overall massive quantity of data available in the corpora, it was important to do a random sample in order to be scientifically sound. However, not every word had 50 available tokens in the corpora so total, the dataset ended up with 1,962 tokens. These tokens were saved in a spreadsheet, where surrounding words (e.g., definite

---

[3] http://davies-linguistics.byu.edu/personal/

[4] This process was automated by a Python script written by Earl Kjar Brown at Brigham Young University.

articles, indefinite articles, possessive adjectives, descriptive adjectives, etc.) were used to determine the gender of anglicisms by noting the gender of those flanking words, and each token was coded as either masculine of feminine. For instance, words preceded by *meu* 'my' POSS PRO MASC, *seu* 'your' POSS PRO MASC, *teu* 'your' POSS PRO MASC , and *nosso* 'our' POSS PRO MASC were classified as masculine while words that followed *minha* 'my' POSS PRO FEM, *sua* 'your' POSS PRO FEM, *tua* 'your' POSS PRO FEM, and *nossa* 'our' POSS PRO FEM were categorized as feminine.

As the tokens were coded according to their gender, some examples had no indication of whether the anglicism was feminine or masculine—that is, the word was neither followed nor preceded by any type of adjective or determiner typically associated with any gender. Therefore, these tokens were excluded from the study. Other tokens that were excluded were when the anglicism was used as a proper noun. Some of the anglicisms, such as *iPhone*, *LAN House*, and *Messenger*, are already proper nouns and are thus proper in every instance of their use. However, when a loanword that is not typically a proper noun was used as a proper noun, it was omitted from the study. This is because when the anglicism is combined with another word to become a proper noun, or is just a proper noun on its own, it is typically because it is being used as the name of a different entity, such as a store or an event, and is no longer referring to what it represents in the language as an English loanword. Once the anglicism becomes part of the title of something, its gender could change, thus skewing the results of this study. For example, with the English spelling of the word *pudding*, most tokens have no distinction for either masculine or feminine. However, 4 tokens were masculine and none were feminine. Four tokens were excluded as proper nouns. One such proper noun is preceded by the word *da* which would ordinarily indicate feminine gender. In this particular instance of the proper noun, *pudding* is part of the name of a street called "Pudding Lane," and the name of this street is what receives

feminine gender, not the word *pudding* itself. Therefore, if the gender of proper nouns were to be included in this study, it is likely that the results would include the gender assignment of other things besides what the loanwords themselves are representing. For this purpose, proper nouns were excluded (when the anglicism itself is not considered a proper noun).

3.4.1 Variability Score

After each token was coded for its grammatical gender, a variability score was created for each individual loanword using the data found for the two genders of that word. For example, the English spelling of the word 'club' had 31 masculine tokens and 4 feminine tokens. Those 35 tokens that indicated a gender were used to create a variability percentage for the word 'club.' The same process was repeated for each anglicism in the dataset, with the particular quantities of gender-indicating masculine and feminine tokens for each word. The scores for each anglicism are included in Table 3 in section 4.1 of this study.

The main focus of this study is to see how attestation in dictionaries and orthographic adaptation influence the variation of gender assignment to anglicisms in Brazilian Portuguese. This research also seeks to determine whether the gender of the anglicism's cognate or most common calque significantly influences the gender most commonly assigned to the loanwords. Finally, this study determines whether the majority gender for each anglicism as indicated by corpus data accurately represents the gender that the dictionaries say should be assigned. In other words, it seeks to find out if what is actually happening is what dictionaries indicate should be happening.

To summarize the process, a list of 40 anglicisms were chosen from various online lists of English loanwords used in Portuguese. For each of the 40 words, 50 tokens were scraped from two corpora in the Corpus do Português. Not every one of the 40 words had 50 tokens available

from the corpora, so the total number of tokens scraped was 1,962. The tokens that indicated a gender for each individual loanword were used to create a variability percentage for each loanword, resulting in only 37 tokens to run statistical tests to determine factors that influence variability.

3.5 Excluded Anglicisms

Three of the 40 English loanwords had to be omitted from the analysis because of their lack of analyzable data. These words are *black*, *sandwich*, and *blêiser* 'blazer'.

The word *black* was eliminated from the study because its appearance in the corpora did not reflect the anticipated meaning. Some of the lists used in order to make the list of anglicisms indicated that *black* is an English loanword, and the author's personal experience with this anglicism is that it is used to describe the typical hairstyle of many people of sub-Saharan African descent, similar to what American English speakers might call an 'afro.' However, the corpus data do not represent this meaning, so it had to be omitted. Most of the examples of the word *black* were proper nouns, such as "Black Friday," "Black Sabbath," and "Black Bloc."

The word *sandwich* (the English spelling) was removed from the study because none of its examples in the 50 tokens from the corpora represented masculine or feminine gender. Out of 50 samples, 48 were proper nouns and the other two had no distinction for either of the two genders.

Finally, the word *blêiser* (the Portuguese spelling of English 'blazer') was removed from the study because no samples with this particular spelling appeared in either of the two corpora. Multiple online lists of English loanwords that aided in the decision of which anglicisms would be analyzed in this study indicated that *blêiser* is the appropriate orthographically adapted

spelling of the word; however, neither corpus returned any results, so this word also had to be omitted.

4.1 Research Question 1 - Attestation in Portuguese Dictionaries

The first factor is how attestation in Portuguese dictionaries influences variation in gender assignment. The research question for this factor is, "What effect does an anglicism's level of attestation in Portuguese dictionaries have on the consistency of gender assignment?" All 40 anglicisms were searched for in four Portuguese dictionaries—*Priberam*[5], *Michaellis*[6], *Infopédia*[7], and *Dicio*[8]. The following anglicisms appear in all four dictionaries: *DVD*, *feedback*, *Internet*, *kit*, *look*, *notebook*, *outdoor*, *selfie*, *blazer*, *clube*, *filme*, *gol*, *piquenique*, *pudim*, *sanduíche*, *telefone*, *vídeo*, and *uísque*. The following English loanwords appear in three of the four dictionaries: *freezer*, *hobby*, and *blêiser*. None of the English loanwords in this dataset appear in exactly two of the four dictionaries. The following anglicisms appear in only one of the four dictionaries: *bike*, *fake news*, *iPhone*, *LAN house*, *ticket*, and *whiskey*. Finally, the following anglicisms do not appear in any of the four dictionaries: *black*, *like*, *messenger*, *pet shop*, *print*, *club*, *film*, *goal*, *picnic*, *pudding*, *sandwich*, *telephone*, and *video*. A table showing which loanwords appear in which dictionaries can be found in Appendix 7.3.

Fifty random tokens (25 from each corpus) were collected for most of the anglicisms in the research, as some loanwords did not have 50 tokens available in the corpora. The word *clube* ended up only having 49 overall tokens, while *fake news* only had the 25 from the NOW corpus. The lack of occurrences of *fake news* in the Web/Dialects corpus is presumably because the phrase predates the rise in frequent usage of the phrase in the media, likely influenced by US

---

[5] https://dicionario.priberam.org/
[6] https://michaelis.uol.com.br/
[7] https://www.infopedia.pt/
[8] https://www.dicio.com.br/

President Donald Trump's 2016 election. *Outdoor* only had 38 tokens, and the English spelling

of *pudding* only had 37. The word *pet shop* was searched and scraped three different ways—*pet

shop* (as two separate words), *petshop* (as all one word), and *pet-shop* (as a hyphenated word)—

in order to have an accurate representation of the gender assignment, in case there is discrepancy

amongst Portuguese speakers in how the phrase is spelled. As two separate words, *pet shop* had

the full 50 tokens, but as all one word, *petshop*, there were only 38 tokens. There were even less

with the hyphenated version, *pet-shop*, with 18. These quantities were combined to find the

variability score. At the time this data were collected in January of 2020, the Python script only

gathered six examples of the word *selfie* from the Web/Dialects corpus. Therefore, 31 total

samples of the word were collected (as the expected 25 were scraped from the NOW corpus).

The original six samples that were collected from the Web/Dialects corpus were numbers 1-3

and 5-7. Thus, samples 8 through 26 out of 29 total samples from Brazil were then added later in

order to complete the 25 random samples from each corpus. The anglicisms *telephone* and

whiskey only had a total of 49 and 41 tokens, respectively. The rest of the 40 anglicisms had the

full 50 tokens.

Not every token in the random sample indicated a gender. Unless the loanword was

preceded or followed by an adjective or an determiner, there was no way of telling which gender

was assigned to the anglicism. In order to determine the variation of the gender assignment, only

the tokens whose surrounding words indicated either masculine or feminine gender were

included. As previously mentioned, tokens in which the anglicisms were used as proper nouns

were also excluded.

In order to determine the percentage of variation, the percentage of masculine

occurrences and the percentage of feminine occurrences first had to be found. The percentage of

masculine occurrences was found by going word-by-word and dividing the number of masculine tokens by the total number of tokens that indicated either gender and then multiplying that number by 100. The total number of tokens that indicated a gender differed from word to word. The same process was repeated for the percentage of feminine occurrences, though this percentage could also be found by subtracting the masculine percentage from 100. The formulas to find the percentage of masculine tokens and the percentage of feminine tokens are written as follows:

$$ratio\ masculine$$
$$= total\ masculine\ tokens/(total\ femine\ tokens$$
$$+\ total\ masculine\ tokens)$$
$$ratio\ feminine$$
$$= total\ feminine\ tokens/(total\ feminine\ tokens$$
$$+\ total\ masculine\ tokens)$$

After the percentages of masculine and feminine occurrences were found, the percentage of agreement between the two genders could be found. Since there are only two options for a possible gender, masculine or feminine, the most differed possibility in variation between the two is 50%. In other words, the most varied that the data representation could be for each word is 50-50. If half of all tokens that indicate a gender are masculine and the other half are feminine, this would result in 50% variation. If there were three possible genders, the most varied that the data could be is 33.33%. Thus, the way to find the percentage of agreement between the two genders is by subtracting either the masculine or feminine percentage from 0.5, taking the absolute value of that number so that it does not matter whether it was the masculine or feminine percentage, and then dividing the total by 0.5 in order to put the number on a scale from 0 to 100

rather than 0 to 50. The total could also be multiplied by two rather than divided by 0.5 to achieve the same result. The formula that finds the percentage of agreement between the two genders can be written as follows:

$$percent\ agreement\ =\ 100\ *\ (1 - abs(0.5 - ratio\ masculine)/0.5)$$

Because variation is essentially the opposite of agreement, the percentage of variation is just the inverse of the percentage of agreement. This total can be found by subtracting the entire previous formula from 1. The highest possible variation percentage would be 100% if the responses were split 50-50. Complete agreement (there are tokens of only one gender) would be 0% variation. Table 3 below shows the percentage of masculine and feminine tokens as well as the percentage of variability for each anglicism.

Table 3: Variability Percentages

Anglicism	Masc. Tokens	% Masc.	Fem. Tokens	% Fem.	% Variability
Bike	0	0.00%	9	100.00%	0.00%
Black	N/A	N/A	N/A	N/A	N/A
Fake news	0	0.00%	11	100.00%	0.00%
iPhone	31	100.00%	0	0.00%	0.00%
LAN House	0	0.00%	40	100.00%	0.00%
Like	2	100.00%	0	0.00%	0.00%
Messenger	32	100.00%	0	0.00%	0.00%
Pet shop	57	78.08%	16	21.92%	43.84%
Print	13	92.86%	1	7.14%	14.29%

Ticket	25	78.13%	7	21.88%	43.75%
DVD	33	100.00%	0	0.00%	0.00%
Feedback	22	100.00%	0	0.00%	0.00%
Freezer	45	100.00%	0	0.00%	0.00%
Hobby	30	100.00%	0	0.00%	0.00%
Internet	0	0.00%	39	100.00%	0.00%
Kit	36	100.00%	0	0.00%	0.00%
Look	39	100.00%	0	0.00%	0.00%
Notebook	35	100.00%	0	0.00%	0.00%
Outdoor	24	100.00%	0	0.00%	0.00%
Selfie	0	0.00%	10	100.00%	0.00%
Blazer	21	80.77%	5	19.23%	38.46%
Club	31	88.57%	4	11.43%	22.86%
Film	1	100.00%	0	0.00%	0.00%
Goal	27	100.00%	0	0.00%	0.00%
Picnic	20	86.96%	3	13.04%	26.09%
Pudding	4	100.00%	0	0.00%	0.00%
Sandwich	N/A	N/A	N/A	N/A	N/A
Telephone	28	100.00%	0	0.00%	0.00%
Video	26	100.00%	0	0.00%	0.00%
Whiskey	4	100.00%	0	0.00%	0.00%

Blêiser	N/A	N/A	N/A	N/A	N/A
Clube	43	97.73%	1	2.27%	4.55%
Filme	47	100.00%	0	0.00%	0.00%
Gol	24	100.00%	0	0.00%	0.00%
Piquenique	25	100.00%	0	0.00%	0.00%
Pudim	23	100.00%	0	0.00%	0.00%
Sanduíche	30	96.77%	1	3.23%	6.45%
Telefone	28	100.00%	0	0.00%	0.00%
Vídeo	38	100.00%	0	0.00%	0.00%
Uísque	12	92.31%	1	7.69%	15.38%

These figures, along with the number of how many dictionaries in which each anglicism appears, were put into a table. Figure 1 below visualizes the relationship between the number of dictionaries in which anglicisms are attested and the percentage of variation of gender assignment. The percentage of variation was placed on the y-axis while the amount of dictionaries in which the words appeared was placed on the x-axis.

Figure 1: Percentage Variability of Gender Assignment of Anglicisms in Brazilian Portuguese by Attestation in Dictionaries
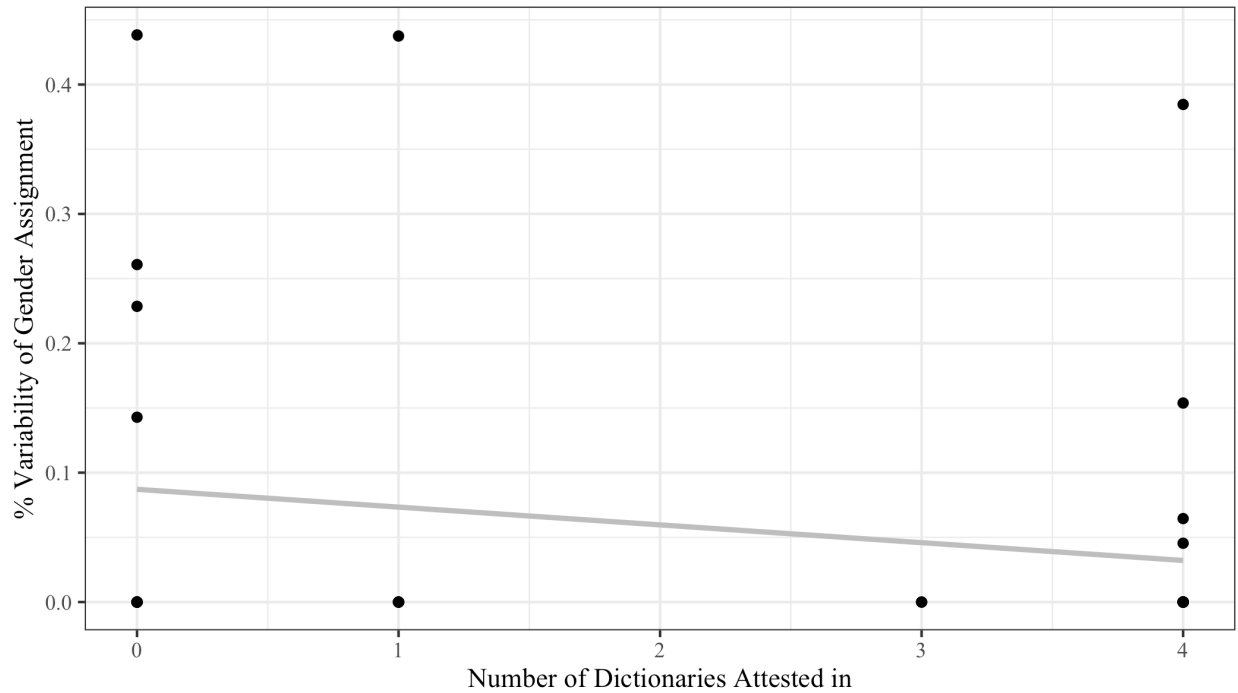
A negative correlation was found between the percentage of variation and the number of Portuguese dictionaries in which anglicisms were attested: the more dictionaries in which the English loanwords were listed, the lower the percentage of variation. This line shows a linear regression between percent variability and number of dictionaries in which anglicisms are found. A linear regression model was fitted and a p-value of 0.222 was found for the number of dictionaries, as seen in Table 4 below.

Table 4: Linear Regression of Percentage Variability as a Function of Number of Dictionaries in which Anglicisms are Attested

| Variable	Coefficient	Std. Error	t-value	p-value

Num. Dict.	-0.014	0.0111	-1.243	0.2219

Multiple R^2: 0.0412; Adjusted R^2: 0.0145

4.2 Research Question 2 - Orthographic Adaptation

Some loanwords' orthographies have been changed to reflect Portuguese orthographic norms. In the particular dataset used for this study, ten of the 40 English words have both a Portuguese spelling and an English spelling. Both spellings are included in the list. In other words, half of the 40 words in the list of anglicisms comprise the same ten words that are spelled two different ways. The two different spellings of each of the ten words were analyzed in the corpora and compared to those words that maintain original English orthography in order to determine whether more variation occurs amongst loanwords spelled with the original English orthography or with the adapted Portuguese orthography. The research question regarding this factor is, "Do anglicisms that have changed orthographically towards Portuguese orthography show less variation in gender assignment than anglicisms that have not undergone an orthographic change?" The anglicisms in the dataset for this study that have undergone an orthographic change are *blêiser* (blazer), *clube* (club), *filme* (film), *gol* (goal), *piquenique* (picnic), *pudim* (pudding), *sanduíche* (sandwich), *telefone* (telephone), *vídeo* (video), and *uísque* (whiskey). Some of these alterations are clearly more drastic than others, with some differences being only one letter or an accent (e.g., *clube*, *filme*, *vídeo*), and others having an entirely different spelling (e.g., *piquenique*, *uísque*). In order to determine whether the orthographically altered anglicisms show less variation in their gender assignment than words that have maintained their original English spelling, each of the loanwords' variability scores were put into

32

a table along with whether the word was altered orthographically and their variation percentages. The table showing the variation percentage can be found in section 4.1. With that data, a Wilcoxon rank sum with continuity correction test was run and a scatterplot was created in order to visualize the relationship between a loanword's orthography and the percentage of variation of gender assignment. This scatterplot can be seen in Figure 2 below. The percentage of variation was placed on the y-axis while the loanword orthography was placed on the x-axis, separated into English orthography on the left side and Portuguese orthography on the right.

Figure 2: Percentage Variability of Gender Assignment of Anglicisms in Brazilian Portuguese by Anglicism Orthography
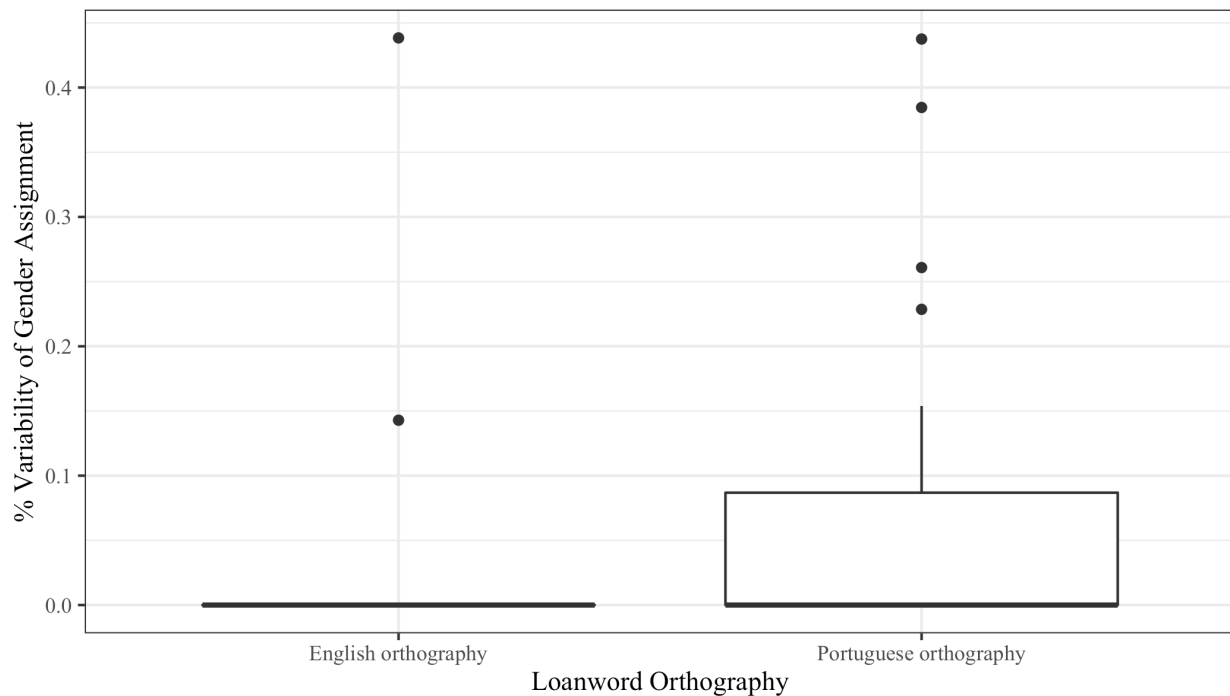
The box on the right side of the chart indicates that there is more variation in gender assignment amongst the Portuguese spellings (e.g., *piquenique*, *clube*, etc.) of anglicisms than there are with the English spellings (e.g., *picnic*, *club*, etc.). A Wilcoxon rank sum with

continuity correction test was also run in order to determine a p-value. The p-value came back as not significant, at a p-value of 0.112 (W =139).

4.3 Research Question 3 - Cognate or Calque Gender

As previously mentioned in the Review of Literature, Geerts' hierarchy (1996, as cited by Franco, et. al, 2018) claims that cognate status, semantic group, suffix, and translation equivalent, in that order, are the most salient factors when it comes to gender assignment in Dutch. Thus, according to Geerts (1996, as cited by Franco, et. al, 2018), cognate is most relevant as it pertains to gender assignment, but the translation equivalent's gender is used if none of the previous factors are present. The anglicisms in this study did not fit into a semantic category that is typically assigned a particular gender in Portuguese, nor did any of them have a suffix that generally indicates a gender. Therefore, only the gender of cognates and translation equivalents were considered. Translation equivalent here is referred to as 'calque.'

However, in their study of anglicisms used in the Spanish spoken in the southwestern United States, Clegg and Waltermire (2009) wrote that the most notable factors are biological gender if the loanword refers to an animate object, and then terminal phoneme if it does not. If the anglicism does not have a terminal phoneme that indicates a particular gender, English loanwords were assigned masculine gender by default. Synonyms, or translation equivalent, do not play a significant role in the determination of gender assignment according to Poplack and Sankoff (1982) as cited by Clegg and Waltermire (2009). What this part of the research sought to determine is the relative importance of this factor in this dataset of anglicisms in a sample of two registers in Brazilian Portuguese. The research question for this factor is "Does the gender used for anglicisms by the majority of people as represented in the corpus data agree with the gender of the cognate or most frequent calque?" It is not a matter of whether Geerts (1996, as cited by

Franco, et. al, 2018) was correct in saying that calque does in fact influence gender assignment to anglicisms when no other factors are present or whether Clegg and Waltermire (2009) were accurate in saying that a loanword's calque does not have a significant impact on its gender assignment. Both studies correctly represent their respective languages—Geerts (1996, as cited by Franco, et. al, 2018) found that calque does influence English loanword gender assignment in Dutch when no other factors are present, and Clegg and Waltermire (2009) quote Poplack and Sankoff (1982) saying that calque does not influence English loanword gender assignment in Spanish, and their research shows this to be true. Again, the grammatical gendering system of Dutch is different than that of Spanish in that Dutch has neuter and common gender while Spanish has masculine and feminine. Spanish grammatical gender is also highly correlated with word final '-o' or '-a,' while in Dutch, the terminal phoneme doesn't correlate with neuter or common gender. Thus, the purpose of this research question is to determine whether or not an anglicisms' cognate or calque is a significant indication of gender assignment in this sample of anglicisms used in two registers of Brazilian Portuguese.

To analyze whether the cognate or calque gender has a significant influence on the assigned anglicism gender, each anglicism from the dataset was inputted into a table along with the most common gender as indicated by the corpora, the gender of the cognate or most common calque, and whether or not those two genders were the same. For instance, the anglicism *bike*, which means 'bicycle,' has the cognate *bicicleta* in Portuguese. The corpus data indicate that the most commonly assigned gender for *bike* is feminine, and its cognate, *bicicleta* is also feminine. Thus, the two genders are the same. One example of the anglicism's corpus-indicated gender and calque gender not matching is with the word *print*. This word is what Brazilian Portuguese speakers use to describe a 'screenshot' and the corpus data designate it as masculine. However,

the calque for this word is *captura de tela* and is feminine. Therefore, the two genders are not the same. The table with all of these details for every anglicism can be found in Appendix 7.5. For example, Using this data, a scatter plot was created and a Wilcoxon rank sum with continuity correction test was run. The p-value of 0.433 (W=182) is not statistically significant. The scatter plot is shown below in Figure 3.

Figure 3: Percentage Variability of Gender Assignment of Anglicisms in Brazilian Portuguese by Cognate or Calque Gender
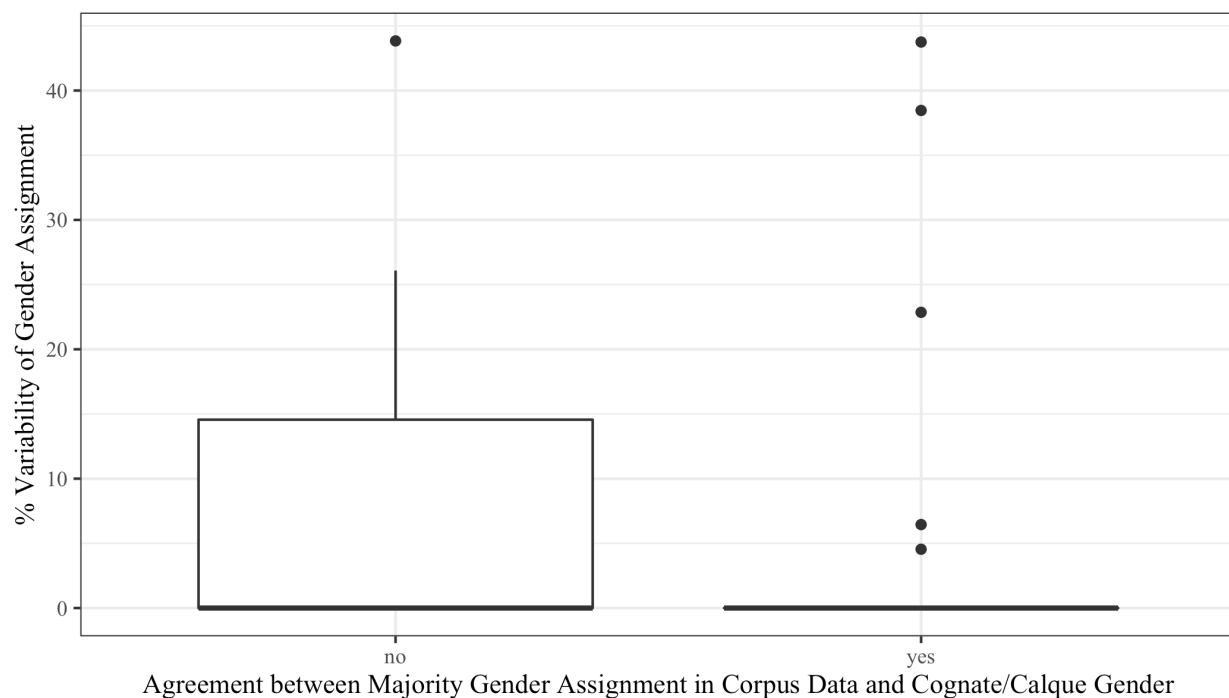
4.4 Research Question 4 - Dictionary Gender Versus Actual Usage

The research in this section seeks to answer whether the genders that Brazilian Portuguese speakers typically give to anglicisms are the same genders that dictionaries indicate should be in use. The specific question is, "For the anglicisms that appear in at least one

Portuguese dictionary, do the findings from the corpus data accurately depict what the dictionaries indicate that the gender should be?" Several English loanwords did not appear in any of the Portuguese dictionaries. These words are: *black*, *like*, *Messenger*, *pet shop*, *print*, *club*, *film*, *goal*, *picnic*, *pudding*, *sandwich*, *telephone*, and *video*. The word *LAN house* appears in one dictionary but its gender was not specified. While the Portuguese spelling of *blêiser* appears with masculine gender in three out of four Portuguese dictionaries, this spelling was not found in either of the two corpora, so no data were available for it.

Therefore, these 15 words do not assist in the answering of this question, but the remaining 25 words were analyzed as to whether their majority gender in the corpora matches the gender that the majority of dictionaries indicate should be the gender. Table 5 below shows the following information:

- the number of dictionaries in which an anglicism is attested (excluding the words listed above)

- the gender as indicated by the majority of dictionaries

- the number of tokens that indicate a gender in the random sample collected by the corpora

- how many of those tokens indicate masculine gender

- how many of those tokens indicate feminine gender

- the percentage of masculine tokens (the percentage of feminine tokens can easily be found by subtracting this number from 100)

Table 5: Dictionary Gender and Corpus Data

Anglicism	Number of dictionaries in which	Gender as indicated by	Number of tokens that indicate a	Number of masc. tokens	Number of fem. tokens	% masculine

	word is attested (out of 4)	majority of dictionaries	gender in corpora			
Bike	1	Feminine	9	0	9	0%
Blazer	4	Masculine	26	21	5	80.77%
Clube	4	Masculine	44	43	1	97.73%
DVD	4	Masculine	33	33	0	100%
Fake news	1	Feminine	11	0	11	0%
Feedback	4	Masculine	22	22	0	100%
Filme	4	Masculine	47	47	0	100%
Freezer	3	Masculine	45	45	0	100%
Gol	4	Masculine	24	24	0	100%
Hobby	3	Masculine	30	30	0	100%
Internet	4	Feminine	39	0	39	0%
iPhone	1	Masculine	31	31	0	100%
Kit	4	Masculine	36	36	0	100%
Look	4	Masculine	39	39	0	100%
Notebook	4	Masculine	35	35	0	100%
Outdoor	4	Masculine	24	24	0	100%
Piquenique	4	Masculine	25	25	0	100%
Pudim	4	Masculine	23	23	0	100%
Sanduíche	4	Masculine	31	30	1	96.77%
Selfie	4	Feminine	10	0	10	0%
Telefone	4	Masculine	28	28	0	100%

Ticket	1	Masculine	32	25	7	78.13%
Uísque	4	Masculine	13	12	1	92.31%
Vídeo	4	Masculine	38	38	0	100%
Whiskey	1	Masculine	4	4	0	100%

Some of these words have important issues to note:

The word *selfie* appears in all four Portuguese dictionaries, but has a bit of a discrepancy as to which gender it should be assigned. Two of the dictionaries (*Priberam* and *Dicio)* say it could be either masculine or feminine while the other two (*Michaellis* and *Infopédia*) listed it as only feminine. However, all four dictionaries label it with the potential to be feminine, while only two indicate that it could be either one. Two of 50 samples are masculine while 11 are feminine, thus deeming this word as feminine because of the majority.

The anglicism *sanduíche* (spelled in Portuguese) also had a bit of a discrepancy as well in the dictionaries' assignment of gender. The *Priberam* dictionary said that this word is masculine in Brazil and feminine in Portugal. The *Michaellis* and *Dicio* dictionaries labeled this word as strictly masculine, while the dictionary *Infopédia*, a dictionary based on Porgtuguese from Portugal, labeled it as feminine.

Of the words whose English and Portuguese orthographies were analyzed in this study, the only ones whose English and Portuguese spellings both appeared in any of the dictionaries were *blazer* and *whiskey*. The English spelling of *blazer* actually appeared in more dictionaries than the Portuguese spelling, which is an anomaly. All four dictionaries indicated this anglicism as masculine. The corpus data agree with this, as 21 tokens are masculine while 5 are feminine. The English spelling of *whiskey* only appears in one of the dictionaries, and it shows up as

masculine as well. The word *whiskey* has 4 masculine tokens in the corpus, with no feminine tokens.

5 DISCUSSION

5.1 Research Question 1 - Attestation in Portuguese Dictionaries

The research question for this section was, "What effect does an anglicism's level of attestation in Portuguese dictionaries have on the consistency of gender assignment?" It was found that the more dictionaries in which an anglicism is attested, the less variability exists within the assignment of gender. Although the p-value (0.222) is not significant, there is a negative trend between the variation score and the number of dictionaries.

The lack of significance is likely due to the fact that the statistical values are based on only 37 data points. This is because at most 50 tokens (but sometimes even less) of each loanword were used to create a single variation score for each anglicism, and those scores were used in the plots and statistical tests. With more loanword types, it is likely that a significant effect would have been found.

Although the p-value is not significant, the negative trend was to be expected. It is logical to think that the more dictionaries in which an English loanword is attested, the more established the word has become in Portuguese, and the more agreement exists among which gender to use in the data from the two corpora used in this study. The anglicisms that are not highly attested, (those that are not listed in any dictionaries or that only appear in one dictionary) are likely less established in the language and thus have a higher discrepancy of which gender is used.

5.2 Research Question 2 - Orthographic Adaptation

The research question for this section was, "Do anglicisms that have changed orthographically towards Portuguese orthography show less variation in gender assignment than anglicisms that have not undergone an orthographic change?" It was found that there is more

variation in the gender assignment of words that have been changed orthographically. This result does not show the expected trend. It was hypothesized that the words with Portuguese orthography would show less variation in gender assignment, but they ended up showing more. It seems that orthographic adaptation to Portuguese would be evidence that the word has been appropriated into the language and is therefore more likely to have only one gender. Differently, newly borrowed words that are still represented with English orthography are more likely to still be in a state of flux as to which gender it has.

Even though the p-value (0.112) is not significant, this could be due to the fact that this statistical analysis only had 40 tokens total (37 excluding the three omitted words), which was caused by the statistically sound method of conducting a random sample, as previously explained. It could also be due to the fact that there were only ten orthographically changed loanwords (nine words after *blêiser* was excluded: *clube*, *filme*, *gol*, *piquenique*, *pudim*, *sanduíche*, *telefone*, *vídeo*, *uísque*) while there were 30 anglicisms with English spelling (28 after *black* and *sandwich* were excluded). Perhaps a more balanced representation of words that were adapted orthographically towards Portuguese and those that were not would achieve different results.

5.3 Research Question 3 - Cognate or Calque Gender

The research question for this section was, "Does the gender used for anglicisms by the majority of people as represented in the corpus data agree with the gender of the cognate or most frequent calque?" The purpose of this question was to discover whether or not the cognate or calque of loanwords has a significant effect on loanwords in Brazilian Portuguese. Geerts (1996, as cited by Franco, et. al, 2018) found that cognates and calques influence gender assignment in Dutch, but Clegg and Waltermire (2009) confirmed Poplack and Sankoff's (1982) findings that

synonyms in the host language do not impact gender assignment in Spanish. Thus, this study sought to determine the effect of cognate or calque gender on anglicism gender.

A Wilcoxon rank sum test with continuity correction was run, comparing the loanwords' gender most commonly assigned in the corpora, and whether or not that gender was the same as the gender of the cognate or most common calque (e.g., whether the masculine *hobby* matches the masculine calque for *hobby*, which is 'passatempo'). A Wilcoxon rank correlation was run rather than the more common Pearson's r correlation test because the data are not normally distributed, as seen in Shapiro-Wilk normality tests. The p-value of 0.433 (W=182) is not statistically significant. However, there is a trend that shows there is more variation in the gender of loanwords when the gender of a cognate or calque differs from the gender that the majority of tokens of a word type have in the two corpora. This supports the idea that the cognate or calque influences the gender of anglicisms. Again, the lack of significance could be due to the fact that the random sample consists of so few tokens that indicate a gender, with only 37 individual cognate or calque genders.

5.4 Research Question 4 - Dictionary Gender Versus Actual Usage

What this section sought to find was the answer to the question, "For the anglicisms that appear in at least one Portuguese dictionary, do the findings from the corpus data accurately depict what the dictionaries indicate that the gender should be?" In other words, is what the dictionary or dictionaries indicate should be happening actually happening amongst real users of Brazilian Portuguese?

The answer to this question is a clear yes. For every single one of the 26 loanwords that appear in at least one of the four online Portuguese dictionaries used in this study, the majority gender as represented in the corpus is the same gender that appears most commonly in the

dictionaries. That is not to say that every single token used the same gender as the dictionary-assigned gender, but the most frequently used gender for each word was the same as the one that most dictionaries had assigned. This shows that these words are well-integrated into the language.

Even some of the words' entries in the dictionaries indicate the potential for variability. Of the words in this dataset, *selfie* and *sanduíche* are the only two anglicisms that have differences in the dictionaries as to which genders should be assigned. For instance, two of the dictionaries label the word *selfie* as strictly feminine, while the other two indicate that it could either be feminine or masculine. Two dictionaries indicate that *selfie* could either be masculine or feminine, but since the other two indicate it as strictly feminine, it was decided that this word would be coded as having feminine gender according to the dictionaries. One dictionary (*Infopedia*), which is based on Portuguese from Portugal, indicates that the word *sanduíche* should be feminine. This is clarified when another dictionary (*Priberam*) lists the word *sanduíche* as being masculine in Brazil and feminine in Portugal. However, since this study deals with Brazilian Portuguese, and because three of the four dictionaries indicated this anglicism as masculine (as *Priberam* indicated that it could go either way), this loanword's majority dictionary gender was labeled as masculine. The discrepancy in gender assignment in the dictionaries for the Portuguese spelling of *sanduíche* and *selfie* indicates the potential for variation. The variability in the gender assignment of dictionary entries shows that variation in the corpus results is not unexpected. In the two corpora analyzed here, of the 50 tokens for the word *sanduíche*, 30 had masculine gender.

5.5 Identification of Most Relevant Factor

Similar to how Geerts (1996, as cited by Franco, et. al, 2018) created a hierarchy for the factors that influence gender assignment to English loanwords in Dutch, after this data were collected and analyzed, the author sought to discover which factor—the gender of cognates and calques or the dictionary gender—most strongly impacted the gender assignment to anglicisms in Brazilian Portuguese. However, after comparing the genders of the 26 words that appeared in at least one dictionary to the genders of the cognates or most common calques, it became apparent that the dictionary genders and the calque genders were mirror images of each other. The results for all 26 of the words were identical, so there is no way to determine which of the factors is more significant. Therefore, neither the gender of the translation equivalent nor the gender indicated by the dictionary are more important than the other in the assignment of nominal gender to anglicisms in Brazilian Portuguese. There may be other factors that are more or less significant, but none that were included in this study.

6 CONCLUSION

6.1 Answers to Research Questions

This section will sum up the findings, limitations, and implications of this research. Table 6 below revisits the research questions.

Table 6: Research Questions Revisited

1. What effect does an anglicism's level of attestation in Portuguese dictionaries have on the consistency of gender assignment?
2. Do anglicisms that have changed orthographically towards Portuguese orthography show less variation in gender assignment than anglicisms that have not undergone an orthographic change?
3. Does the gender used for anglicisms by the majority of people as represented in the corpus data agree with the gender of the cognate or most frequent calque?
4. For the anglicisms that appear in at least one Portuguese dictionary, do the findings from the corpus data accurately depict what the dictionaries indicate that the gender should be?

For research question 1, it was hypothesized that there would be a negative correlation between variability and anglicism attestation in Portuguese dictionaries. The p-value of 0.222 is not significant, but there is a trend showing that the more dictionaries in which an English loanword is attested, the lower the variation percentage tends to be. For research question 2, it was hypothesized that there would be more variation amidst gender assignment of anglicisms with original English spelling than anglicisms with altered Portuguese spelling. This hypothesis was wrong, as a box and whiskers plot of the corpus data indicates that more variability actually exists among words with Portuguese orthography. However, the p-value (0.112), once again, is not significant. For research question 3, it was hypothesized that the gender of the cognate or calque would have a significant effect on the gender of the anglicism. This hypothesis was, once

again, wrong, as the p-value is 0.433. Though the p-value is not significant, these results show a trend that there is greater variation in the gender assigned to anglicisms that have a different gender as indicated by the corpus than the gender of the cognate or most frequent calque. Finally, for research question 4, it was hypothesized that the corpus data showing the genders that actual Portuguese speakers are using would accurately represent the genders that the dictionaries indicate. This hypothesis was correct, because for each one of the loanwords that appeared in at least one of the dictionaries, its corpus-indicated gender matched the majority gender as indicated by the corpus data.

Though the results were not significant, it is important to note the small number of tokens in this study, intentionally chosen to ensure statistical soundness and manageability. The 1,962 tokens were used to create a variability score for each individual anglicism, resulting in only 37 tokens to analyze variability. With a larger dataset including more of the ample corpus data, it is possible that results could be significant.

6.2 Limitations

Though this research was planned and carried out with thoughtful care, its limitations should be considered. Arguably the most obvious limitation is the small amount of tokens, due to the data collection method of random sampling. Though nearly 2,000 individual tokens were collected from the corpora (and many more were available in the corpora), these tokens were simply coded and used to determine the variability of gender assignment for each individual type. After having to exclude three words due to lack of analyzable data in the randomly selected tokens, only loanwords remained. The necessity to exclude certain loanwords due to their lack of data with the anticipated usage (i.e., *black*), or their lack of data altogether (i.e., *blêiser*), was certainly another limitation. That's not to say that the corpora didn't provide sufficient data—the

tokens that were randomly selected did not fit the expectations of the research. Another limitation includes lack of gender-indicating tokens appearing for every randomly selected token in the corpora to help identify which gender is being assigned. Finally, though the Corpus do Português is a phenomenal corpus, and both the NOW corpus and the Web/Dialects corpus that were used for this study have 1 billion words or more each, all of the language therein is Internet-based, so no books or spoken language are accounted for. Therefore, while the Internet-based written language in the Web/Dialects and NOW corpora may correspond with spoken language, the lack of representation of all types of language in the corpora is one limitation. This is no fault of these particular corpora; due to the nature of corpora themselves, it would not be possible to account for every potential genre in a given language in order to analyze all the instances of loanword use.

6.3 Ideas for Future Research

Since the corpora consist of a great deal of data that was not used in this analysis, another data collection process could be employed to quickly accrue larger amounts of data. For example, if a proxy search were to be made using search terms such as "o | no | um | seu | meu | este | esse" and "a | na | uma | sua | minha | esta | essa" to precede the anglicism and a search were to be conducted in that way, a vastly higher number of tokens could be collected. This would not give results for each and every instance in the corpus because it's only using seven determiners that occur one word to the left of the target word, when in reality some of these words occur two words to the left and many adjectives that indicate gender appear one word to the right , but it would certainly give a higher number of tokens that this study had. However, though this data collection method would result in larger amounts of data, it would not achieve the statistic goal of having a random sample. In order to be scientifically sound, every member of the population

has to have the same probability of appearing in the random sample, which the proxy searches do not afford.

Another limitation that could potentially be avoided in future research is that of having few tokens that indicate a gender. From the random sample of 50, sometimes very few tokens indicated either masculine or feminine gender. There are enough data in the corpora to get 50 samples of tokens that do indicate a gender, so a more robust analysis could come about by continuing to collect tokens from a random sample until there are 50 tokens that indicate one gender or the other. Further research could be done with more types (i.e., other anglicisms used frequently in Brazilian Portuguese) or with the same number of types but just a different dataset to see if the results are consistent.

Many of the English loanwords in this dataset deal with technology (i.e., *iPhone*, *LAN house*, *like*, *Messenger*, *DVD*, *print*, *Internet*, *selfie*, *telefone*, *vídeo*). Identifying semantic categories into which anglicisms can be sorted may be beneficial, and may show that semantic category has some significance, as it does in Dutch (Geerts, 1996 as cited by Franco, et. al, 2018). Other factors besides semantic category could also be analyzed as to how they influence gender assignment and variability in gender assignment. These factors may include animacy (biological gender), amount of time in the language (when the loanword was first introduced), and frequency in usage.

It appears that the masculine gender could be the unmarked or default gender, while deviations from this are clearly tied to the gender of the cognate or calque when it appears as feminine. Examples of this include *bicicleta* 'bike', *casa* 'house', and *notícias* 'news' in the instances of *bike*, *LAN house*, and *fake news*. It seems that cultural borrowings, or concepts that are new to the culture, such as *Messenger* and *iPhone* receive the unmarked gender, that is, the

masculine gender. Another potential future analysis could look at the significance or the correlation between whether a loanword is a core or cultural borrowing as defined by Myers-Scotton (2002) and its gender, to identify whether the masculine gender truly is the unmarked, default gender that all ideas or concepts that are new to the language automatically receive.

Another study could also look at these same questions in Portuguese from Portugal or another Portuguese speaking country. If it were for a country other than Portugal, Mozambique, or Angola, (like Cape Verde, for instance), then a different corpus would have to be used or even created. Though studies like this one have been done for several other languages (German, Dutch, Italian, Spanish, etc.), there are undoubtedly other languages that have not been analyzed yet with regards to variability in the gender assignment of English loanwords. A study duplicating this one could be carried out but using other corpora; a corpus consisting of strictly Twitter data or data gathered elsewhere could be made.

6.4 Implications

This research has pedagogical implications. For Portuguese learners, the findings of this study show that an anglicism's gender is very often influenced by its cognate or calque—thus, if a non-native speaker would like to use an English loanword but isn't sure which gender to use with it, the speaker has a good chance of getting the gender correct by just associating the gender of the cognate or calque.

Another implication of this research is that variability exists among gender assignment to English loanwords among native speakers of Portuguese. Some words, such as those that are less established in the language, tend to have more variability than others. This being said, using either gender with an anglicism is likely acceptable, since the data from the corpora show that there are discrepancies in the ways that native speakers assign gender. Though some have been

established in dictionaries, the ones that haven't been could go either way with gender assignment. This implication shows that with these less attested words, it essentially doesn't matter which gender is assigned. Though no anglicism's variability was split evenly with gender distribution, many of the less established ones had masculine and feminine tokens.

In conclusion, anglicisms of many different kinds exist in Brazilian Portuguese. Whether borrowed for creativity, innovation, prestige, or to name a new concept, English loanwords are constantly being used in Portuguese. Some of these words are much more established and thus vary less in their gender assignment, while some words that are not as established vary more in their gender assignment. More tokens in the random sample would likely help the results achieve statistical significance, but the random sample itself provides statistical solidity, so the trends are reliable.

7 APPENDIX

7.1 Grammatical Glosses

Following is the interpretation of what each grammatical description used in the text means:

NOUN MASC - Masculine noun
NOUN FEM - Feminine noun
DEF ART MASC - Definite masculine article
DEF ART FEM - Definite feminine article
INDEF ART MASC - Indefinite masculine article
INDEF ART FEM - Indefinite feminine article
PLUR DEF ART MASC - Definite plural masculine article
PLUR DEF ART FEM - Definite plural feminine article
PLUR INDEF ART MASC - Indefinite plural masculine article
PLUR INDEF ART FEM - Indefinite plural feminine article
ADJ MASC - Masculine adjective
ADJ FEM - Feminine adjective
PLUR ADJ MASC - Plural masculine adjective
PLUR ADJ FEM - Plural feminine adjective
POSS PRO MASC - Masculine possessive pronoun
POSS PRO FEM - Feminine possessive pronoun
PLUR POSS PRO MASC - Masuline plural[9] possessive pronoun
PLUR POSS PRO FEM - Feminine plural possessive pronoun
PREP ART CONT MASC - Masculine preposition-article contraction masculine
PREP ART CONT FEM - Feminine preposition-article contraction feminine
PLUR PREP ART CONT MASC - Masculine reposition-plural article contraction
PLUR PREP ART CONT FEM - Feminine preposition-plural article contraction

---

[9] As in, what is being possessed is plural. Who is possessing it could be singular or plural.

7.2 Collocates

Following is a list of collocates that help in the identification and distinction as either masculine or feminine:

Masculine:
- ❖ *meu* 'my' POSS PRO MASC
- ❖ *meus* 'my' PLUR POSS PRO MASC
- ❖ *teu* 'your' POSS PRO MASC
- ❖ *teus* 'your' PLUR POSS PRO MASC
- ❖ *seu* 'your' POSS PRO MASC
- ❖ *seus* 'your' PLUR POSS PRO MASC
- ❖ *nosso* 'our' POSS PRO MASC
- ❖ *nossos* 'our' PLUR POSS PRO MASC
- ❖ *pelo* 'for the' 'by the' 'through the' PREP ART CONT MASC
- ❖ *pelos* 'for the' 'by the' 'through the' PLUR PREP ART CONT MASC
- ❖ *no* 'in the' 'on the' 'at the' PREP ART CONT MASC
- ❖ *nos* 'in the' 'on the' 'at the' PLUR PREP ART CONT MASC
- ❖ *do* 'of the' 'from the' PREP ART CONT MASC
- ❖ *dos* 'of the' 'from the' PLUR PREP ART CONT MASC
- ❖ *o* 'the' DEF ART MASC
- ❖ *os* 'the' PLUR DEF ART MASC
- ❖ *um* 'a' 'an' INDEF ART MASC
- ❖ *uns* 'some' PLUR ADJ MASC
- ❖ *algum* 'some' ADJ MASC
- ❖ *alguns* 'some' PLUR ADJ MASC
- ❖ adjectives that end in o(s) (e.g., *barato(s)* 'cheap')

Feminine:
- ❖ *minha* 'my' POSS PRO FEM
- ❖ *minhas* 'my' PLUR POSS PRO FEM
- ❖ *tua* 'your' POSS PRO FEM
- ❖ *tuas* 'your' PLUR POSS PRO FEM
- ❖ *sua* 'your' POSS PRO FEM
- ❖ *sua* 'your' PLUR POSS PRO FEM
- ❖ *nossa* 'our' POSS PRO FEM
- ❖ *nossas* 'our' PLUR POSS PRO FEM

- *pela* 'for the' 'by the' 'through the' PREP ART CONT FEM
- *pelas* 'for the' 'by the' 'through the' PLUR PREP ART CONT FEM
- *na* 'in the' 'on the' 'at the' PREP ART CONT FEM
- *nas* 'in the' 'on the' 'at the' PLUR PREP ART CONT FEM
- *da* 'of the' 'from the' PREP ART CONT FEM
- *das* 'of the' 'from the' PLUR PREP ART CONT FEM
- *a* 'the' DEF ART FEM
- *as* 'the' PLUR DEF ART FEM
- *uma* 'a' 'an' INDEF ART FEM
- *umas* 'some' PLUR ADJ FEM
- *alguma* 'some' ADJ FEM
- *algumas* 'some' PLUR ADJ FEM
- adjectives that end in a(s) (e.g., *nova(s)* 'new')

7.3 Definitions

Below is a brief description of what each anglicism in this dataset means as it is used in Portuguese. Starred entries have different meanings in Portuguese than they do in English.

- ❖ *Black (noun): used to describe the hairstyle of many black people, similar to what American English speakers might call an 'afro'
- ❖ *Outdoor (noun): a billboard
- ❖ *Pudim (noun): the Portuguese spelling of a gelatin-based dessert; this dessert is not the same as American pudding; it is more similar to the Mexican dessert flan
- ❖ Bike (noun): a bicycle
- ❖ Blazer (noun): the English spelling of a professional-looking lightweight overcoat
- ❖ Blêiser (noun): the Portuguese spelling of a professional-looking lightweight overcoat
- ❖ Club (noun): the English spelling of a particular group of people who share a unifying characteristic; can also refer to a place of recreation that requires a membership to enter
- ❖ Clube (noun): the Portuguese spelling of a particular group of people who share a unifying characteristic; can also refer to a place of recreation that requires a membership to enter
- ❖ DVD (noun): Digital Video Disc (the gloss for this in Portuguese would be *disco de vídeo digital*, so the acronym would still be DVD if it was translated)
- ❖ Fake news (noun): disinformation; news that is not true
- ❖ Feedback (noun): constructive criticism given in response to an action or performance
- ❖ Film (noun): the English spelling of a movie
- ❖ Filme (noun): the Portuguese spelling of a movie
- ❖ Freezer (noun): a compartment of a refrigerator (or sometimes separate) that keeps food and liquids frozen
- ❖ Goal (noun): the English spelling of how a point is scored in soccer
- ❖ Gol (noun): the Portuguese spelling of how a point is scored in soccer
- ❖ Hobby (noun): an activity done for leisure in one's free time
- ❖ Internet (noun): the World Wide Web
- ❖ iPhone (noun): a smartphone made by Apple
- ❖ Kit (noun): a group of items arranged in a packet that serve a similar purpose
- ❖ LAN House (noun): similar to a "cybercafe;" a computer establishment where people can pay to use the Internet; "LAN" stands for "Local Area Network," in English; it is part of the word in Portuguese just as it is in English
- ❖ Like (noun): a reaction on social media sites, such as Facebook, Twitter, YouTube, and Instagram
- ❖ Look (noun): someone's personal style

- ❖ Messenger (noun): an instant messaging and video calling app in connection with Facebook
- ❖ Notebook (noun): a laptop; a personal and portable computer
- ❖ Pet shop (noun): a store where one can purchase animals and pet supplies; also spelled *petshop* and *pet-shop*
- ❖ Picnic (noun): the English spelling of when a meal is taken to an outdoor space and eaten there
- ❖ Piquenique (noun): the Portuguese spelling of when a meal is taken to an outdoor space and eaten there
- ❖ Print (noun): a screen capture on a cellphone or computer
- ❖ Pudding (noun): the English spelling of a creamy gelatin-based dessert
- ❖ Sanduíche (noun): the Portuguese spelling of
- ❖ Sandwich (noun): the English spelling of a type of food that consists of two slices of bread with some type of filling, generally meat, cheese, and vegetables, or with nut butter and fruit jam
- ❖ Selfie (noun): a picture taken of oneself
- ❖ Telefone (noun): the Portuguese spelling of a technological device used to call and speak to people who are not nearby
- ❖ Telephone (noun): the English spelling of a technological device used to call and speak to people who are not nearby
- ❖ Ticket (noun): a slip of paper granting admission to a show, concert, movie, or public transportation
- ❖ Uísque (noun): the Portuguese spelling of a particular alcoholic beverage
- ❖ Video (noun): the English spelling of a visual recording
- ❖ Vídeo (noun): the Portuguese spelling of a visual recording
- ❖ Whiskey (noun): the English spelling of a particular alcoholic beverage

7.4 Table of Attestation in Portuguese Dictionaries

Anglicism	*Priberam*	*Michaellis*	*Infopédia*	*Dicio*
Bike	-	feminine	-	-
Black	-	-	-	-
Blazer	masculine	masculine	masculine	masculine
Blêiser	masculine	masculine	masculine	-
Club	-	-	-	-
Clube	masculine	masculine	masculine	masculine
DVD	masculine	masculine	masculine	masculine
Fake news	-	-	-	feminine
Feedback	masculine	masculine	masculine	masculine
Film	-	-	-	-
Filme	masculine	masculine	masculine	masculine
Freezer	masculine	masculine	-	masculine
Goal	-	-	-	-
Gol	masculine	masculine	masculine	masculine
Hobby	masculine	-	masculine	masculine
Internet	feminine	feminine	feminine	feminine

iPhone	-	masculine	-	-
Kit	masculine	masculine	masculine	masculine
LAN House	-	listed with no gender	-	-
Like	-	-	-	
Look	masculine	masculine	masculine	masculine
Messenger	-	-	-	-
Notebook	masculine	masculine	masculine	masculine
Outdoor	masculine	masculine	masculine	masculine
Pet shop	-	-	-	-
Picnic	-	-	-	-
Piquenique	masculine	masculine	masculine	masculine
Print	-	-	-	-
Pudding	-	-	-	-
Pudim	masculine	masculine	masculine	masculine
Sanduíche	masculine in Brazil	masculine	feminine in Portugal	masculine
Sandwich	-	-	-	-
Selfie	feminine or masculine	feminine	feminine	feminine or masculine

Telefone	masculine	masculine	masculine	masculine
Telephone	-	-	-	-
Ticket	-	-	-	masculine
Uísque	masculine	masculine	masculine	masculine
Video	-	-	-	-
Vídeo	masculine	masculine	masculine	masculine
Whiskey	-	-	masculine	-

7.5 Table of Cognate/Calque Gender

Anglicism	Gender from corpora	Cognate or calque	Gender of cognate or calque	Corpora and calque gender the same?
Bike	feminine	bicicleta	feminine	yes
Black	N/A	N/A	N/A	N/A
Blazer	masculine	casaco especial	masculine	yes
Blêiser	N/A	N/A	N/A	N/A
Club	masculine	grupo especial	masculine	yes
Clube	masculine	grupo especial	masculine	yes
DVD	masculine	disco digital de vídeo	masculine	yes
Fake news	feminine	notícias falsas	feminine	yes
Feedback	masculine	comentários sobre uma situação	masculine	yes
Film	masculine	apresentação audiovisual	feminine	no
Filme	masculine	apresentação audiovisual	feminine	no
Freezer	masculine	congelador	masculine	yes
Goal	masculine	um ponto no futebol	masculine	yes
Gol	masculine	um ponto no futebol	masculine	yes
Hobby	masculine	passatempo	masculine	yes
Internet	feminine	rede	feminine	yes
iPhone	masculine	telefone celular da marca de Apple	masculine	yes
Kit	masculine	conjunto de peças	masculine	yes

LAN House	feminine	casa de rede	feminine	yes
Like	masculine	curtida	feminine	no
Look	masculine	estilo próprio	masculine	yes
Messenger	masculine	mensageiro	masculine	yes
Notebook	masculine	computador portátil	masculine	yes
Outdoor	masculine	painel grande que serve como propaganda	masculine	yes
Pet shop	masculine	loja de animais	feminine	no
Picnic	masculine	refeição no ar livre	feminine	no
Piquenique	masculine	refeição no ar livre	feminine	no
Print	masculine	captura de tela	feminine	no
Pudding	masculine	sobremesa feita de gelatina	feminine	no
Pudim	masculine	sobremesa feita de gelatina	feminine	no
Sanduíche	masculine	tipo de alimento que consiste em duas fatias de pão com algo no meio	masculine	yes
Sandwich		tipo de alimento que consiste em duas fatias de pão com algo no meio	masculine	yes
Selfie	feminine	foto tirada de si mesmo	feminine	yes
Telefone	masculine	aparelho usado para ligar	masculine	yes
Telephone	masculine	aparelho usado para ligar	masculine	yes
Ticket	masculine	bilhete	masculine	yes
Uísque	masculine	bebida alcoólica	feminine	no
Video	masculine	uma gravação visual	feminine	no

Vídeo	masculine	uma gravação visual	feminine	no
Whiskey	masculine	bebida alcoólica	feminine	no

8. References

Anderman, G. M., & Rogers, M. (2005). *In and Out of English: For Better, for Worse?* Bristol: Multilingual Matters.

Clegg, J. H., & Waltermire, M. (2009). Gender Assignment to English-Origin Nouns in the Spanish of the Southwestern United States. *Southwest Journal of Linguistics*, *28*(1), 1–14.

Davies, Mark. (2016-) Corpus do Português: One billion words, 4 countries. Available online at http://www.corpusdoportugues.org/web-dial/. (Web / Dialects)

Figueiredo, E. H. D. D. (2010). To borrow or not to borrow: the use of English loanwords as slang on websites in Brazilian Portuguese. *English Today*, *26*(4), 5–12. doi: 10.1017/s0266078410000301

Franco, K., Zenner, E., & Speelman, D. (2018). Let's Agree to Disagree. (Variation in) the Assignment of Gender to Nominal Anglicisms in Dutch. *Journal of Germanic Linguistics*, *30*(1), 43–87.

Friedrich, P. (2002). English in advertising and brand naming: sociolinguistic considerations and the case of Brazil. *English Today*, *18*(3), 21–28. doi: 10.1017/s0266078402003048

Gilbert, G. G. (1965). English Loanwords in the German of Fredericksburg, Texas. *American Speech*, *40*(2), 102. doi: 10.2307/453716

Gomes, L., & Lopes, J. G. P. (2011). Measuring Spelling Similarity for Cognate Identification. *Progress in Artificial Intelligence Lecture Notes in Computer Science*, 624–633. doi: 10.1007/978-3-642-24769-9_45

Haspelmath, M., & Tadmor, U. (2009). *Loanwords in the world's languages: a comparative handbook*. New York, NY: De Gruyter Mouton.

Haugen, Einar (1950), 'The Analysis of Linguistic Borrowing,' *Language*, 26:210-31.

Hoffer, B. L. (2002). Language Borrowing and Language Diffusion: An Overview. *Intercultural Communication Studies*.

Myers-Scotton, C. (2002). *Contact linguistics: bilingual encounters and grammatical outcomes*. New York: Oxford University Press.

Oncins-Martínez, J. L. (2009). Towards a corpus-Based Analysis of Anglicisms in Spanish: A Case Study. *International Journal of English Studies*, 115–132.

Onysko, A., Callies, M., & Ogiermann, E. (2013). Gender variation of anglicisms in German: The influence of cognitive factors and regional varieties. *Poznań Studies in Contemporary Linguistics*, *49*, 103–136. doi: 10.1515/psicl-2013-0004

Perini Mário A. (2002). *Modern Portuguese: a reference grammar*. New Haven, CT: Yale University Press.

Poplack, S. (1993). Variation theory and language contact. American dialect research, ed. By

    Dennis Preston, 251-286. Amsterdam: John Benjamins.

Rabeno, A., & Repetti, L. (1997). Gender Assignment of English Loan Words in American

    Varieties of Italian. *American Speech*, *72*(4), 373–379. doi: 10.2307/455494

Ribeiro dos Santos, A. de C. (2016).

Sá Pereira, M. de L. (1948). *Brazilian Portuguese grammar*. Boston: Heath.

Sicherl, E. (1999) *The English Element in Contemporary Standard Slovene. Phonological,*

    Morphological and Semantic Aspects. Ljubljana: University of Ljubljana.

Thomason, Sarah & Kaufman, Terrence (1988), *Language Contact, Creolization and Genetic*

    Linguistics. Berkeley, CA: Univ. of California Press.