

MEMORIAL UNIVERSITY
OF NEWFOUNDLAND
OCCASIONAL PAPERS
IN LINGUISTICS
Volume 2

*Proceedings of the 33rd Atlantic Provinces Linguistics Association
Memorial University of Newfoundland*

Sara Johanson, Evan Hazenberg and Suzanne Power
Editors

Paul De Decker
Faculty Supervisor

MEMORIAL UNIVERSITY OF NEWFOUNDLAND
ST. JOHN'S
2011

Memorial University of Newfoundland
Occasional Student Papers in Linguistics
c/o Department of Linguistics
Memorial University of Newfoundland
St. John's, NL Canada A1B 3X9
<http://www.mun.ca/linguist>

Copyright © 2011

Anglicisms: The Numbers Behind the Words

JESSE HARRIS & WALCIR CARDOSO

1. Introduction

Over the decades, linguists around the world have studied the infiltration and usage of one language in another language. Some view this type of cross-linguistic influence as interference (see Mougeon & Beniak 1991, Weinreich 1974), while others see borrowings as a source of enrichment to the language (see Guiraud 1965, Le Prat 1980, Picone 1996). Regardless of the various sentiments towards the phenomenon, given the long trail of borrowings throughout linguistic history, it is safe to assume that languages will continue to borrow from each other well into the future.

The borrowing of English words in particular has interested many scholars, especially speakers of French. Historically speaking, Wise (1997) points out that the borrowing of English words into French dates back to the 1600s with a significant increase in anglicisms starting by the end of the 18th century. De Ullmann (1947) even cites a very small number of English borrowings before the 1600s with one instance (alderman) first appearing in French in 1363.

Eventually, English borrowings began to create worry amongst certain French-speaking communities ultimately resulting in legal measures taken to protect the language from English “contamination”. According to Nadeau and Barlow (2006), Quebec spearheaded language protection after World War II in 1959, which eventually lead to the Loi 101: Charter of the French Language in 1976. In fact, it was France that then modeled their well-known language protection policies after the movement in Quebec. Previous language protection attempts in France had proved largely ineffective up to that point due to the lack of belief by French authorities that there truly was a threat posed by the English language. However, French literary critic René Étimble’s (1964) “Parlez-vous franglais?” opened many French speakers’ eyes to the prevalence of English in France spawning the founding of the Délégation générale à la langue française et aux langues de France (DGLFLF) in the 1990’s, a committee charged with the protection of the French language in France .

And so, has the effort put forth by the governments of Quebec and France to guard against outside English language elements influenced the actual everyday usage of English words and phrases in the French language today? What kind of role do words like “feeling” and “tripper” play in the informal, unmonitored speech of a typical Francophone? In order to answer these questions, one must first consider the actual amounts and percentages of anglicisms in the

language as a whole. In this study I will seek to address the question of anglicism frequency and distribution in French. The next section will provide an overview of pertinent research and findings on the topic of anglicisms in French followed by a terminology section on anglicisms in general and relevant background. Next, three research questions will be presented followed by a detailed section on variable selection and corpus design. A description of data manipulation and analysis will follow, and finally the paper will conclude with a summary as well as a brief discussion on possible contributions of the proposed research.

2. Previous Research

In order to investigate more specifically the nature of borrowings across languages, this study focuses on borrowings between two languages in particular: English as the donor language and French as the recipient language. The political “hotness” of the anglicism topic in both Quebec and France generated a host of research and critiques from linguists in both regions particularly in the 1970’s-1990’s (Nadeau & Barlow 2006). Book chapters and journal articles dedicated to the topic of anglicisms, however, have been largely descriptive, devoting energy primarily to explaining the nature of and reasons for English borrowings in French. Less frequent have been research dedicated to the pure numbers. The following section briefly reviews the empirical studies that have already addressed, from varied perspectives, the frequency of anglicisms in French.

The number of studies devoted exclusively to quantifying anglicisms in French is somewhat sparse. Authors Forgue (1986), Théoret (1991) and Mareschal (1992) all took interest in this linguistic phenomenon, considering a number of common variables important to the investigation of anglicisms in French.

One important variable considered by each author was the variety of French to study. As different varieties of French are inevitably attached to different cultural, political, and historical events and norms, language variety notwithstanding, the use and frequency of anglicisms could conceivably vary from region to region. In Forgue’s (1986) study, for example, the author undertook the task of identifying anglicisms in the variety of French spoken in France. Théoret (1991) also investigated a single variety of French from various regions of the province of Quebec in Canada (Estrie, Montreal, Quebec, and Saguenay-Lac St-Jean). Finally, in Mareschal (1992), the author presented results from research on anglicisms in four different francophone regions: Belgium, France, Quebec, and Switzerland.

The number and the source of the words studied varied between authors as well. Forgue (1986) analyzed just under fifteen million words (13,700,000 words), and his corpus consisted of daily articles from the French newspaper *Le Monde* in 1977. These newspaper articles provided data for only one language mode: written language. The potentially problematic nature of this fact, as admitted by Forgue himself, is that *Le Monde* was a newspaper generally seen as “elitist” employing a highly formal register of language. The likelihood that the language in this newspaper (including anglicization) was a representative sample of current day language at the time of the study is slim. In Théoret (1991), the corpus was considerably smaller than that of Forgue (1986). The Sherbrooke Corpus was a corpus of one million words made up fifty percent of “spontaneous oral” language and fifty percent “non-spontaneous oral” language (i.e., language “written to be spoken”: folklore, theatre, radio broadcasts, soap operas, monologues, etc.). Like Forgue, the Sherbrooke corpus represented only one mode of language, comprised only of

spoken (not written) language. Although the author does not disclose the total size of the corpora, Mareschal (1992), as with Forgue (1986), also gathered data for her research from newspapers during the late 1970s and early 1980s. Unlike Forgue, the author used newspapers accessible to the general public.

On the other hand, all three authors employed a similar approach to quantifying the linguistic variable. This was a dual calculation approach which took into consideration anglicism word tokens as well as word types. For example, for the token calculation Théoret (1991) conducted a simple frequency count of the total number of individual anglicism occurrences versus the total number of word occurrences in the corpus. Additionally, the author derived an anglicism type count by tallying the total number of anglicism word types (n = 699) in the corpus and dividing by the total number of French word types (n = 11,327). Forgue (1986) and Mareschal (1992) also considered both anglicism tokens and types in their analyses. The obvious importance of taking both word tokens and word types into consideration in any frequency analysis will be discussed further in Section 5, below.

Of course, the results and their implications proved the most interesting comparisons between these three authors. The results of Forgue's (1986) research produced a total of 8,200 anglicism tokens, translating into roughly 0.60% of the collected corpus. Moreover, when counting types (n = 680), this figure dropped to 0.04%. Théoret's (1991) frequency count unveiled 2,861 anglicism tokens in the corpus of one million French words (0.28%), however for the analysis of type of anglicism, the result was a figure of 6% indicating that six percent of the total distinct word types (English or French) in the corpus were comprised of anglicisms. The results of Mareschal's (1992) analysis uncovered 1,801 total tokens of anglicisms and 904 total distinct anglicism word types. Unique from Forgue and Théoret, Mareschal compared the distribution percentages of anglicisms between French-speaking regions rather than measuring the total percentage of anglicisms in the French language.

The relevant information and results from the above literature review are summarized in Table 1.

Table 1: Summary of previous anglicism research

Author	Forgue (1986)	Théoret (1991)	Mareschal (1992)
Corpus size	15 M	1 M	?
Source	Written (Le Monde daily, 1977)	Oral (Sherbrooke corpus)	Written (Daily newspapers)
Variety	France	Quebec	Belgium, France, Quebec, Switzerland
Results	0.6% / 0.04% (tokens/ types)	0.28% / 6.0% (tokens/ types)	n = 1,801 / 904 (tokens/ types)

The design and results of these studies are revealing in many ways regarding the variables adopted. However, to date, there still exists a gap in the literature on anglicism frequency in French. The most salient inadequacy is the timeliness of the data. Not only were the primary studies on anglicism frequency carried out over two decades ago, but also the data in these studies were already somewhat obsolete at the time of publication. Another shortcoming involves the inconsistency in variables studied across the literature. While some authors choose

to look at only one mode (written or spoken language) across several regions (language varieties), other authors focus only on one language mode in one language variety.

We are not aware of any study to date that incorporates current data into research on both language modes across more than one language variety. Consequently, drawing from method and design present and/or lacking in the research reviewed here, the objective of the current study is to collect and analyze anglicism frequency data from up-to-date sources in both written and spoken language across two distinct varieties of French.

3. Terminology: What's in a Name?

To date, there has been a substantial amount of research on the topic of anglicisms and their integration in various languages. General definitions for anglicisms have been embedded within literature and empirical studies, and even dictionary forewords and introductions provide explanations on the topic (see the *Petit Robert* 2005, the *Dictionnaire de Français Plus* 1988, Villers 1988, and Rey-DeBove & Gagnon 1980). Of the available literature, however, only a small number of books and articles on the topic have attempted a clear definition. What is more, of the works that actually define “anglicism”, very few have gone beyond a general explanation to consider the depth or breadth of the variety of important linguistic factors contributing to the identification of an anglicism (with some exceptions: see Mareschal 1992, and Picone 1996 for comprehensive categorizations of anglicisms). As a result, confusion in the naming terminology (and presumably the identification) of various borrowings from English ensues.

An exploratory investigation and synthesis of the term “anglicism” and its various definitions in the literature has revealed approximately six different categories of anglicisms in the French language: *wholesale anglicisms*, *direct translations*, *semantic anglicisms*, *hybrids*, *French inventions and modifications*, and *morphological anglicisms*. The next sections will focus more specifically on the nuances and complexities of these six different kinds of anglicisms.¹

3.1 Wholesale Anglicisms

Wholesale anglicisms (or *Intact/quasi-intact borrowing*, *whole/partial borrowing*, *conscious borrowing*, *direct loan*, and *Frenchified anglicisms*), are a type of anglicism that undergoes (virtually) no change from English to French (see Bouchard 1999, Forgue 1986, Grigg 1997, Guiraud 1965, Mareschal 1992, Rifelj 1996, Spence 1989, Trescases 1982, Villers 1988). The word or expression in English is identical in both form and meaning to its usage in French. For example the English word “weekend” is used with identical orthography and meaning in French. The same is true for “muffin”, “fair play”, and “bowling”. That is to say that the French language imports an English object as well as its corresponding English meaning.

3.2 Direct Translations

A second type of anglicism is English words, expressions, or ideas directly translated from English (usually morpheme for morpheme) into French (see Forgue 1986, Grigg 1997, Guiraud

¹ Although divided here into distinct and separate categories, it is often the case that anglicisms produced in both speech and writing overlap and prove to be composed of elements from more than one of these categories.

1965, Mareschal 1992, de Ullmann 1947, Wise 1997). These direct translations (most commonly called *calques*, *structural calques*, or *loan translations*) may include “gratte-ciel” for *sky scraper*, “haut parleur” for *loud speaker* or even “bienvenue” for *welcome*, the short form of *you’re welcome*.² Grigg (1997) as well as Mareschal (1992) and Picone (1996) also look at the direct translation of English compounds into French. According to the literature, this syntactic process can happen in one of two ways: either the English compound is borrowed into French, taking on the French word order as in “facteur-risque” (for *risk factor*), or the English compound is borrowed into French imitating the English compounding structure (with a left-headed modifier) such as “télérepas” (for *TV dinner*). Thus, direct translation is essentially the borrowing of an English object, its meaning, and occasionally its syntax while leaving behind the English name.

3.3 Semantic Anglicisms

By and large, the most difficult type of anglicism to identify is the semantic anglicism (or *semantic borrowing*, *semantic calque*, *loan shift*, and *semantic imitation*). This type of anglicism constitutes borrowing an English word that is similar to an already existent French word (with a different meaning), and superimposing the English meaning on the French word (see Bouchard 1999, Guiraud 1965, Mareschal 1992, Picone 1996, Rifelj 1996, de Ullmann 1947, Villers 1988, Wise 1997). The most common (and many times the only) example of this type of anglicism in the literature is the French verb “réaliser” being incorrectly employed with the English meaning *to realize*. False cognates like these are the most susceptible kinds of words to this kind of anglicization. In French, the traditional usage of the verb “réaliser” means *to fulfill* or *to achieve*. However, when “réaliser” undergoes semantic anglicisation, it takes on the English definition/meaning of *to realize* (which already exists as “se rendre compte de” in French).³ Because of the subtlety in meaning change as well as the natural difficulty caused by false cognates, semantic anglicisms prove difficult to detect and often go unnoticed even by native speakers of French.

3.4 Hybrids

A small category of anglicisms includes the hybrid as discussed by Grigg (1997), Picone (1996) and Trescases (1982). This anglicism type combines existing French elements with borrowed English words. For example, the hybrid “surbooker” is composed of the English verb *to book* as well as the French prefix “sur” meaning *over*. This results in an anglicism meaning *to overbook*.

² An interesting stipulation brought up by Grigg (1997) is that oftentimes, one needs to have certain knowledge of current English culture in order to understand the directly translated French expression. Two examples are “chasseur de têtes” and “parfum du jour”. Despite the fact that the expression uses entirely French words, a person would need to be familiar with the workforce recruitment agents known as *head-hunters*, to understand. Nor would one necessarily understand the concept of *flavor of the day* commonly used to refer to certain individuals who change love interests on a frequent basis.

³ Rey-DeBove (1980) was the one of the only authors we encountered to provide other examples of this kind of anglicism. She points out that the French word “audience” in the sense of a *hearing*, has adopted the English meaning of *audience* like the crowd at a show. In addition, “alternative” in French means *alternate* as in “an alternate spelling”, whereas, when transformed into a semantic anglicism, it uses the English meaning of *alternative* (one of several possible solutions).

At first sight, hybrids resemble direct translations and wholesale anglicisms. However, whereas direct translations and wholesale anglicisms borrow all elements of the English word or expression into French, hybrids borrow an English word as a base and insert existing French elements around the base. Other examples of hybrids in the literature include “en live” and “top modèle”. Ultimately, one could say that a hybrid maintains the same meaning from English to French, and that the form is a mix between a direct translation and a wholesale anglicism.

3.5 French Inventions and Modifications

The next type of anglicism lies at the opposite end of the spectrum from wholesale anglicisms in that there is no transfer of meaning from English to French. A French invention (also known as *false anglicisms*, *anglicisms of the signifier*, *pseudo borrowing*, *pseudo-anglicisms*, and *over anglicization*) is a word based on English elements that adopts a French meaning which is unusual or unknown to Anglophones (see Forgue 1986, Grigg 1997, Guiraud 1965, Mareschal 1992, Picone 1996, Spence 1989, Thogmartin 1984, Trescases 1982, Villers 1988). English words are given a French meaning that is distorted or does not equate to the meaning of the word in English. For example, the word “tennisman” would appear quite familiar to a native English speaker due to the two English elements *tennis* and *man*. Though one may be able to guess the French meaning of this word, *tennis player* remains the conventional term in English. Another less intuitive example of a French invention is the popular word “footing”. Although the English word *foot* and the English morpheme “ing” are both evident, an anglophone may not necessarily know to bring sneakers when invited to go “footing” (*go for a jog*). Finally, a French speaker may talk about “un lifté” who lives down the street. The English word *lift* is perceptible, but in what sense of the word? In fact, “un lifté” refers to a person who has had a *face lift*; surgically speaking they have been *lifted*.

Another group of anglicisms explored mainly by Grigg (1997) involves the influence of French lexis and syntax on English borrowings. One instance of this type of modification occurs through the truncation (shortening) of an English word to make a French word (see also Mareschal 1992). For example, a French person could very well put on a “sweat” (*sweatshirt*), grab their “walk” (*walkman*) and stroll out to “le parking” (the *parking lot*). Although all three of these truncated anglicisms are wholly English words, they do not carry the same meaning as their non-truncated English counterparts.⁴ In sum, French inventions and modifications borrow an English word into French while leaving behind the English meaning.

3.6 Morphological Anglicisms

Up to this point, the most prevalent categories of anglicisms in French relate to meaning and their transfer into French, yet another small category deserves brief mention. Although limited in scope, Grigg (1997), Spence (1989), and Trescases (1982) all attest to the existence of a morphological anglicism category in French. The most commonly cited morphological anglicism involves the suffixation of the English “-ing” morpheme (e.g., “brushing”). Another example of a morphological anglicism is the verb “lifter” where the French infinitive inflection “-er” is

⁴ Some other French modifications include the English word undergoing a grammatical class change in French (*fitness* (n.) in English becomes “faire fitness” (v.) in French), and singularization as with the English plural word *jeans* becoming the singular French word “un jeans”.

suffixed onto the English word *lift*. In this way we can regard the relatively rare morphological anglicisms as English words that undergo or cause some type of change word-internally when borrowed into French.

The six categories of anglicisms that have been explained in this section are summarized in Table 2 along with some examples.

Table 2: Anglicism by category

Type	Definition	Example
Wholesale anglicism	A word or expression in English that is (usually) identical in both form and meaning to its usage in French.	- weekend, muffin, bowling, fair play, sweat shirt, look (n.), Hi-Fi, steack, rosbif
Direct translation	Borrowing an English object and its meaning and directly translating it into French.	- gratte-ciel, haut parleur, chasseur de têtes, hors de la loi, effet de serre, nettoyage ethnique, facteur-risque
Semantic anglicism	Borrowing an English word that is similar to an already existent French word (with a different meaning), and applying the English meaning to the French word.	- réaliser (in the English sense of to realize, not to fulfill or to achieve as in French) - audience, alternative
Hybrid	An English base word with added French elements.	- surbooker, en live, top modèle
French invention & modification	Borrowing an English word form into French without borrowing the English meaning.	- tennisman, footing, un lifté/ un transplanté, slip (n.) - walk (walkman), straight pipe, un jeans
Morphological anglicism	The addition of a morpheme that changes the meaning of the word through inflection and/or derivation.	- brushing, forcing, lifter (v.)

4. Research Questions

As exemplified in the above review of previous research, frequency studies have touched on important factors such as anglicism use (tokens and types) in different varieties of French, and anglicisms in written language and in spoken language. However no single study to date has investigated a combination of the factors together in one body of research. The current study seeks to fill a gap in anglicism frequency research by addressing these shortcomings through the consolidation of numerous factors in the production of anglicisms in French. This research will explore three questions specifically:

- (1) Which language variety of French (France vs. Quebec) uses a higher total percentage of anglicisms?

- (2) Which language mode (written vs. spoken) is characterized by a higher frequency of anglicisms?
- (3) What is the effect of a type/token frequency distinction in analyzing anglicisms in French?

5. Variable Selection and Corpus Design

The objective of this study is to investigate the percentage of anglicisms in a representative corpus of the French language. Its design is based on a corpus of written and transcribed spoken data collected and compiled specifically for this study so as to take into consideration three major factor categories: *language variety* (French from Quebec, and French from France), *language mode* (written French, and spoken French), and the *token/type* distinction. In order to procure comparable oral and writing samples from French speakers from both France and Quebec, data from two television programs (Star Academy/Star Académie) and text from web logs (blogs) were collected, compiled and analyzed.⁵

5.1 Language Variety

Background. As mentioned above, this study will compare French from two distinct language varieties: French from France and from Quebec. The importance of studying more than one variety of French was highlighted in Mareschal (1992), and thus chosen as a factor for the present study. The reasons for which anglicisms are employed in each of the two language varieties is beyond the scope of this study; however, the authors concur that France and Quebec both borrow anglicisms for different reasons (Bouchard 1999, Martel 1991, Nadeau & Barlow 2003). This topic will be revisited in Section 7, in light of the current findings and their relevance to the research questions posed.

Current design. Because there are different reasons for borrowing and there exist different categories borrowed between users of France French and Quebec French, any study of anglicism frequency requires the separate consideration of the two language varieties. In the case of the current study, language variety will be defined as prototypical French written or spoken in two distinct regions: France and Quebec. The two subcategories France French and Quebec French are defined as the French produced by natives from each region. Natives, in turn, are people not only born in the region, but also residing in (and presumably participating in) the language community at the time of data collection.

5.2 Language Mode

Background. The second main variable under investigation is *language mode*. Research in corpus linguistics and discourse analysis has also revealed that certain linguistic features may vary depending on whether the data are gathered from spoken or written language (Tannen 1982, Chafe 1985, Louwse, McCarthy, McNamara, & Graesser 2004). In most sociolinguistic literature, for example, it is assumed that spoken language is less monitored than written

⁵ One general concern pertains to the importance of random sampling. The scope of this study involves investigating the effects of language variety and language type on anglicisms. This means that certain variables, such as age and gender, will not be accounted for. Yet a certain level of randomness still remains in the sampled television program participants and blog authors. For example, the television program is designed to select participants from various parts of the region. Moreover, the blog data will be completely random other than controlling for the authors' place of birth and residence and topic of discussion. The careful design control allows for this study to be easily replicable in the future by other researchers wishing to investigate the same or even alternate variables.

language. The assumption can be taken one step further to apply to anglicisms in that written language may contain fewer anglicisms if they are seen as “unwanted/undesirable”. On the other hand, as touched upon earlier, previous research has observed different reasons for the use of anglicisms across different language varieties. In regions where English borrowings are viewed as prestigious (i.e., France), the more deliberate monitoring of written language could produce a higher percentage of anglicisms than spoken language due to the author wanting to assert a certain social status.

In the end, the reasons for anglicism use cross-regionally require further investigation, and are beyond the scope of the present study. Chafe and Tannen (1987), in a comprehensive overview of research investigating the differences in written and spoken language, conclude that “different conditions of production as well as different intended uses foster the creation of different kinds of language” (p. 390). It is therefore safe to assume that linguistic borrowings, such as anglicisms in French, may not be exempt from the type of variation that affects spoken and written language modes.

Current design: Written data. In order to procure writing samples from French speakers, text produced (and possibly edited) by authors of various publicly posted web logs (blogs) was collected from Internet websites. For both varieties of French, the content of the blogs remained within the realm of everyday living and family matters (e.g., advice on shopping, job search stories, trouble with spouses and children, etc.).

The question arises, of course, as to whether the authors of these blogs are speakers of the language variety of French from France or from Quebec, a factor which was controlled for in a number of ways. One way of ensuring authorship for the target language variety was by refining Internet searches for blogs to the region-specific domain extensions (i.e., *blogger.fr*, or *blogspot.ca*). Although many French and Quebec authors use blog websites with “.com” extensions, which obscure the authors’ origins, this step was helpful in immediately eliminating many English language blogs. Next, the principle means of confirming the origin and region of residence of the blog authors was via direct personal communication. This communication, in the form of an email, indirectly inquired about the authors’ birth location and region of current residence (either Quebec or France). Once authors had replied, blogs were then chosen based on whether or not replies were consistent with target data. Finally, once the above two methods had been employed, the actual content of the texts, read while reviewing the data, provided valid information about the authors. This last step was especially useful in identifying and eliminating authors who had spent an extended period of time living or traveling in regions where they may have been heavily influenced by the English language (thus potentially skewing the results to reflect a higher frequency of anglicisms in their blogs).

In addition to authorship, the date of the data collected from the blog websites was also taken into consideration, that is, blog postings prior to fall 2008 were not considered for either varieties of French. This ensured that the texts used for data collection were both methodologically equivalent across language varieties as well as quite recent. Finally, the total number of blog sources and words subjected to analysis remained equal across both language varieties.

Current design: Spoken data. The corpus design for the spoken data variable was configured somewhat differently. Data for spoken French for each variety were gathered from the daily lives

of French speakers in two reality television programs: *Star Academy* (in France, www.tf1.fr/star-academy) and *Star Académie* (in Quebec, www.staracademie.ca), two versions of the same program (having the same concept, and format) with the same target audience (for the sake of comparability).⁶

Since the study of anglicisms necessarily involves English words in French, it was important for the sake of internal validity to choose programs that would not be inherently subject to either more or less anglicisms than in typical spontaneous speech. Scripted sitcoms and dramas, for instance, give producers and directors the opportunity to potentially add or delete anglicisms in the script based on the target audience, political standpoints, or other TV network agendas. On the other hand, unscripted reality-format programs analyzed in the present study provided the ideal medium for spontaneous, unmonitored speech and allowed for spontaneous language production in the widest possible range of topics. Moreover, the original series *Star Academy*, produced by European company Endemol, was first aired in France in 2001 (“Star Academy”, 2009) and was adopted in Quebec with the same format and successfully broadcasted as *Star Académie* by 2003. Consequently, the choice of a television program that is not a spin-off of an English concept helped control once again for any unnecessary outside influence from Anglophone culture.

Another important factor in the selection of an appropriate television program is the recency of the data. Previous studies of anglicisms in French are criticized here as using outdated data (e.g., Mareschal 1992, who uses data from the late 1970s). Current data samples are a unique feature of this study and are deemed important as anglicisms are a particularly dynamic part of language (Nadeau & Barlow 2006). Collecting and using recent data assured that the anglicisms found in the data are representative of current day spoken and written anglicisms. Finally, again for comparability sake, language recorded from each of the two television programs represented a similar word count for the two varieties under investigation.

5.3 Tokens and Types

The final factor group of the current study is the distinction between tokens and types of anglicisms. An anglicism token is defined by counting each and every anglicism as a separate occurrence, whereas an anglicism type only counts anglicisms from a new word family (e.g., *blog*, *blogueur*, *blogosphère* constitute three tokens but only one type; the same holds for *weekend*, *weekend*, and *weekend* – three tokens, one type). This token-type aspect of quantification is a particularly important consideration in anglicism frequency counts. Token counts give a sense of density, of whether one particular anglicism is highly used and thus artificially inflating the number of anglicisms, while type counts show the degree of variation/diversity of the different anglicisms in the corpus.

⁶ The difference in orthography between the two television show titles is curious and deserves mention at this juncture. *Star Academy* in France is completely English, a wholesale anglicism, while the Quebec version of the show *Star Académie* appears to attempt a frenchification of the title. Ironically, however, in doing so the title still remains an anglicism but of the direct translation category. Indeed, “Star” aside, in order for the title to be truly French, the title would need to read “Académie des Stars”. Thus, both titles from both varieties of French can be considered anglicisms, but of different categories.

Recall the numbers and percentages discussed in previous studies (see Table 1). The percentage of anglicism occurrences in Forgue (1986) and Mareschal (1992) were all significantly lower for types than for tokens, while the opposite was true for Théoret (1991). It is clear here that if research fails to consider both tokens and types side by side, the reader may only see one aspect of the data. It is therefore imperative for any investigation of anglicism percentages to look not only at the token frequency, but also the frequency of anglicism types in order to get a full understanding of the linguistic feature's true frequency and distribution, and not be misled into drawing under-informed or misguided conclusions from the data.

6. Data Manipulation and Analysis

In order to accurately measure anglicisms, a triangulation of methods and instruments was devised. Firstly, the Microsoft Word spell-check feature in French helped visually locate unrecognized or “misspelled” (i.e., borrowed) words. Once identified as an anglicism, the “find” search feature in Word was employed. This feature proved useful, once a particular anglicism had been identified, in locating any and all other recurring instances of the same anglicism in the corpus.

The next step in identifying anglicisms in the corpus was by referencing the table of definitions in Table 2. A final method for locating anglicisms in the corpus was to consult dictionaries/lists of anglicisms in French (i.e., Forest 2006, Hofler 1982, Laurin 2004, and Rey-DeBove & Gagnon 1980) and cross-reference the information with a French dictionary (Larousse 2004).

The procedure for this study was relatively straightforward. The first step involved the data collection itself. For the *Star Academy/Star Académie* television programs, this involved careful watching and listening to each recorded video file in conjunction with the task of transcribing the files into text. For the blog entries, a certain amount of “clean-up” was necessary in order to render the text into an analyzable corpus. This involved taking out pictures and photos; deleting “non-text” symbols such as smiles and extra punctuation; deleting sounds and verbalized emotions (haha, hihi, lol, pffft!); and deleting outside references or citations from any language (i.e., poems, quotes, movie dialogues, song lyrics, “top 10” lists, recipes, etc.). These measures were taken in order to avoid distorting the data's representative portrayal of anglicisms. Ultimately, for both language types it was necessary to rectify the word count from the data sources to ensure equivalence across language variety.

The second step involved analyzing the data and identifying all instances of anglicisms in the corpus. As described above, anglicisms were identified through a triangulation of methods and instruments: Microsoft Word French spell-check and find feature, a summary chart of definitions from the literature, and dictionaries of anglicisms. This analysis was conducted keeping in mind the three independent variables, language type, language mode (oral/written), and the token/type distinction.

7. Results and Implications

The analysis of the data on both written and spoken anglicism tokens and types in France and Quebec yielded intriguing results. The raw numbers and percentages from the frequency counts,

organized in vertical columns by language variety and language mode, and in horizontal columns by anglicism tokens and types, are presented in Table 3.

Table 3: Anglicism frequency totals

	FF WT	FF SP	FF total	QF WT	QF SP	QF total	Total
Token n=	85	69	154	65	182	247	401
Token %	0.84	0.67	0.75	0.64	1.82	1.23	0.99
Type n=	52	30	82	43	61	104	186
Type %	0.51	0.29	0.40	0.43	0.61	0.52	0.46

Firstly, in looking at the rightmost column “Total” and reading across the first row, one can observe the total number of anglicism tokens in a corpus (written and spoken) of over 40,000 words to be just over four hundred. This translates into just under one percent (second row, 0.99%). As for total number of anglicism types, the number decreases by more than half. That is to say, only 0.46% of the word types in the corpus are anglicism word types (n=186). The interpretation of these totals suggests that the overall percentage of anglicisms in French, when compared with the findings from the previous studies discussed at the start of this paper (see Table 1), has not dramatically increased over the last two decades, as it remains around 1% of the total words analyzed.

Next, the analysis yielded data relevant to anglicism differences across language varieties. The relationship between the total anglicisms in these two language varieties (“FF total” and “QF total” columns) becomes more apparent in Table 4.

Table 4: Anglicisms by language variety

	FF	QF
Token	154 (0.75%)	247 (1.20%)
Type	82 (0.40%)	104 (0.50%)

Recall the first research question posed at the outset of this paper: *Which language variety of French uses a higher total percentage of anglicisms?* In Table 4 it becomes immediately clear that the variety of French spoken and written in Quebec contains more anglicisms than the French from France. The number spread, however is not equal between tokens and types. When looking at the difference in anglicism percentages across the “Token” row, for example, QF contains more anglicisms than FF by almost half a percent (0.45%). On the other hand, however, the difference in the amount of anglicism types used in QF versus the amount in FF is reduced to only a tenth of a percent (0.10%).

These results demonstrate how the Quebec language variety of French uses overall more anglicisms than French in France. However, several caveats are in order regarding additional factors that may have shaped these results. Firstly, due to the difficulty in identification and general rarity of the categories, both morphological anglicisms and semantic anglicisms were omitted from the current data. This resulted in an analysis of only four of the six anglicism categories described earlier in this paper. In order to get the most precise view of anglicism frequency in a language, future research would have to include all six categories.

In addition, the size of the corpus (over forty thousand words total) could have possibly weighted the anglicism occurrences more heavily than in a larger corpus. Although in theory, if a corpus is representative of the population it portrays (which this study has endeavored to be; see previous section on the study’s design), increasing the data set should not alter results. And yet,

the possibility still remains that a higher or lower percentage of anglicisms in one language variety or another could emerge as a result of a larger corpus.

Finally, there remains the fact that certain individual speakers (in the TV data) or authors (in the blog data) may have possessed a certain affinity for or aversion to anglicization. For example, it was observed by one of the current authors (though not formally documented in the data) that one particular speaker from Québec’s *Star Académie* produced a substantially higher number of anglicisms than the other speakers on the show. It is even possible, though harder to observe, that a language user may consciously avoid the use of anglicisms. Regardless of reasons, these individual behaviors would undoubtedly have a smaller effect on the data in a larger corpus.

The answer to the second research question, *Will one language mode yield a higher frequency of anglicisms than the other?*, proves less straight forward than the first (see Table 5). When considering the overall picture of language mode (written versus spoken) it looks as if anglicisms are considerably more prevalent in spoken language (n=251) than in written French text (n=150) when considering word tokens. On the other hand, when looking at anglicism word types, the opposite is true. Ninety-five written anglicism types prove slightly higher than the 91 counted spoken types.

Table 5: Anglicism totals by language mode

	Written	Spoken
Token	150	251
Type	95	91

If stopping here, the implications of these results prove mystifying. And yet, when expanding the view of language mode by adding the variable of language variety, a visible pattern emerges (Table 6).

Table 6: Anglicisms by language mode and variety

	Written	Spoken
FF (token)	85	69
(type)	52	30
QF (token)	65	182
(type)	43	61

In Table 6, it becomes evident (for both tokens and types) that anglicisms are more common in written French in the language variety from France, and that conversely, anglicism use in the Quebec variety of French appears higher in spoken language. The fact that France employs more anglicisms in writing while Quebec uses more in speech could very possibly be attributed to the different reasons each of these language varieties borrows from English. Research on French specifically has found that language users from France use anglicisms in French for different reasons than French language users in Quebec.

Indeed, a vast body of literature has been dedicated to distinguishing the reasons for anglicism use in France and in Quebec (see Timmins 1995, Théoret 1991, Nadeau & Barlow 2003). For example, it has been argued that French speakers from France tend to employ

anglicisms due to their “fondness” of American culture (Timmins 1995) and due to the historical prestige status these English borrowings hold as they have always been associated with higher bourgeoisie social groups and good taste (Bouchard 1999).

Conversely, in Quebec, working-class ex-peasants during the Industrial Revolution began to resent the dominant upper class English society and their novel words/concepts (i.e., *le boss*, *le shop*, *le foreman*, *le drill*, etc.) causing these and other anglicisms to become stigmatized to the point where French speakers looked to replace them with a French form, correct or not (Bouchard 1999, Forest 2006, Timmins 1995). This history, combined with a direct contact with a predominantly English-speaking continent, and continual exposure to English culture through sports, work, and brands, has shaped the nature of and reasons for anglicisms in present-day Quebec French (Nadeau & Barlow 2003, Timmins 1995).

These reasons, of course, are not the only explanations for anglicisms in French, nor have the results of the present study served to definitively propose these explanations as facts. What is clear from the data, however, is that in order to explain why FF contains more anglicisms in written language and why QF has more anglicisms in spoken French, additional research on the relationship between language mode and reasons for borrowing is required.

A final word regarding the third research question posed at the outset of this paper: *Is the distinction between anglicism type frequency and token frequency relevant?* The response to this question manifests itself in the context of the results just discussed. In the first research question pertaining to language variety, for instance, the comparison of token word frequency with type word frequency revealed differences in the distances between the two language types (see Table 4, where we show that the gap in numbers between QF and FF was much smaller for types than for tokens). Moreover, while looking at token frequency gives a general sense of a linguistic feature’s (i.e., anglicisms) pervasiveness and density in the data, type frequency reveals information regarding the diversity (new unique words as opposed to the same word repeating over and over) of that feature in the corpus. That is to say that in the study of anglicism frequency, and in any linguistics frequency study for that matter, it is imperative to take both token and type frequency into consideration. As exemplified in the current data, different results emerge based on whether tokens or types are analyzed. Without simultaneously considering both of these two measures, any results collected from the data will be portraying only part of the picture.

8. Concluding Remarks

This study sought to gather and interpret empirical data regarding the number of anglicisms in a French corpus, in which language mode, language variety, and the distinction between token frequency and type frequency were taken into consideration. Three research questions guided the study, and careful design and procedures were adopted in order to sufficiently answer these questions.

One major contribution of this experiment to cross-linguistic studies is the quantification of a familiar phenomenon (borrowings from English to French) in a methodologically sound way. Théoret (1991), for example, laments the difficulties of speaking about anglicisms in Quebec. Since there is not a large number of objective studies on the subject, people grant themselves the right to speak about and pass definitive judgments on the topic of anglicisms as if

the fact of knowing *how* to speak a language allows them to analyze it in a scientific and impartial manner (p. 79).

The fact that the data collected were recent also brings a substantial amount of validity to the study. This point is especially important to control for when dealing with a dynamic language element such as anglicisms (Clyne 2003). Furthermore, by using reality television shows (spontaneous and unmonitored speech) and blog texts (freely composed by amateur authors), as opposed to edited television programs or newspaper/magazine articles, the corpus compiled and analyzed derives from “authentic” (less monitored) spoken/written contexts.

An additional contribution of this study pertains to the use of both written and spoken data. Many studies to date (see Cerquiglini 1991, Mareschal 1992, Théoret 1991) have investigated anglicisms in either written or spoken corpora, but not both. The significance of incorporating both language modes was discussed earlier in the “Variable Selection and Corpus Design” section of this paper, and the results from the current study have indeed confirmed anglicisms as a linguistic feature susceptible to written language/spoken language differences.

Similarly, another unique aspect of this study is the use of Internet blog data for the written language variety. By using written data from blogs, it was assumed that the level of formality would fall somewhere between formal writing and informal speech. One of the most widely studied variables in sociolinguistics over the past several decades remains the formality of speech. Since Labov’s (1966) famous *fourth floor* study on formality in various situations and contexts, many subsequent studies by researchers in linguistics have sought to uncover the relationship between particular linguistic features and language formality (e.g., Cardoso 1999, 2003, 2007, Lin 2003, Major 2001, 2004). This study has observed a difference between written and spoken language, and yet it remains to be determined whether this difference can be attributed to a potential more careful monitoring of language in either of the two modes investigated.

The goal of this study was to use authentic up-to-date informal writing and speech as tools to investigate the percentages of anglicisms in two varieties of French. Furthermore, compiling a new and unique corpus of current written and spoken data creates an invaluable opportunity for future research on different variables pertaining to anglicisms or any other range of linguistic features in spoken and/or written language. In the end, the question still remains as to at what point the presence of English in everyday French would be considered “contamination” or “infiltration” of the language. Yet, in undertaking such a study, it is hoped that the results will help fill the gaps in the overall body of knowledge in the field of borrowings especially concerning the frequency of anglicism use so that this empirical data can now help substantiate those good-humored finger-pointing bar discussions that claim “your French variety uses more anglicisms than mine”.

References

- BOUCHARD, CHANTAL. 1999. *On n'emprunte qu'aux riches: La valeur sociolinguistique et symbolique des emprunts*. Montreal: Fides.
- CARDOSO, WALCIR. 1999. A quantitative analysis of word-final /r/-deletion in Brazilian Portuguese. *Linguistica Atlantica* 21: 13-52.
- CARDOSO, WALCIR. 2003. Topics in the phonology of Picard. Ph.D. thesis, McGill University. Published by the McGill Working Papers in Linguistics.

- CARDOSO, WALCIR. 2007. The variable development of English word-final stops by Brazilian Portuguese speakers: A stochastic optimality theoretic account. *Language Variation and Change* 19: 1-30.
- CERQUIGLINI, BERNARD. 1991. Le point de vue du français européen. *Actes du colloque sur les anglicismes et leur traitement lexicographique: Communications, discussions et synthèses: Magog du 24 Au 27 Septembre 1991*, ed. by Pierre Martel and Hélène Cajolet-Laganière, 297-301. Quebec: Gouvernement du Quebec, Office de la langue française.
- CHAFE, WALLACE L. 1985. Linguistic differences produced by differences between speaking and writing. *Literacy, language, and learning: The nature and consequences of reading and writing*, ed. by David R. Olson, Nancy Torrance, and Angela Hildyard, 105-123. New York: Cambridge University Press.
- CHAFE, WALLACE L., & DEBORAH TANNEN. 1987. The relation between written and spoken language. *Annual Review of Anthropology*, 16: 383-407.
- CLYNE, MICHAEL G. 2003. *Dynamics of language contact: English and immigrant languages*. Cambridge: Cambridge University Press.
- Dictionnaire du français plus: A l'usage des francophones d'amerique*. 1988. Montreal: Centre éducatif et culturel.
- ETIEMBLE, RENÉ. 1964. *Parlez-vous français?* Paris: Gallimard.
- FOREST, JEAN. 2006. *Les anglicismes de la vie quotidienne des Québécois: Essai*. Montreal: Triptyque.
- FORGUE, GUY J. 1986. English loan words in French today. *Journal of English Linguistics* 19.2: 285-294.
- GRIGG, PETER. 1997. Toubon or not Toubon: The influence of the English language in contemporary France. *English Studies* 78.4: 368-384.
- GUIRAUD, PIERRE. 1965. *Les mots étrangers: Que Sais-Je?* Paris: Presses Universitaires de France.
- HOFER, MANFRED. 1982. *Dictionnaire des anglicismes*. Paris: Larousse.
- LABOV, WILLIAM. 1966. *The social stratification of English in New York City*. Washington: Center for Applied Linguistics.
- LAURIN, JACQUES. 2004. *Les américanismes: 1200 mots ou expressions made in USA*. Montreal: Editions de l'Homme.
- Le petit Larousse illustré: en couleurs*, 100th ed. 2004. Paris: Larousse.
- LE PRAT, GUY. 1980. *Dictionnaire de français: Plus de 850 mots et locutions de langue anglaise couramment utilisés dans les médias, la conversation ou la correspondance française d'aujourd'hui et leur traduction en français*. Paris: Guy Le Prat.
- LIN, YUH-HUEY. 2003. Interphonology variability: sociolinguistic factors affecting L2 simplification strategies. *Applied Linguistics* 24.4: 439-464.
- LOUWERSE, MAX M., PHILIP M. MCCARTHY, DANIELLE S. MCNAMARA, and ARTHUR C. GRAESSE. 2004. Variation in language and cohesion across written and spoken registers. *Proceedings of the twenty-sixth annual conference of the Cognitive Science Society*, ed. by Kenneth Forbus, Dedre Gentner and Terry Regier, 843-848. Mahwah, NJ: Erlbaum.
- MAJOR, ROY. 2001. Linguistic explanations for second language phonological systems. [Chapter 2]. *Foreign Accent. The Ontogeny and Phylogeny of Second Language Phonology*. New Jersey: Lawrence Erlbaum Associates.
- MAJOR, ROY. 2004. Gender and stylistic variation in second language phonology. *Language Variation and change* 16.3: 169-188.
- MARESCHAL, GENEVIÈVE. 1992. L'influence comparée de l'anglais sur le français dans différentes aires géographiques francophones. *Journal of the Canadian Association of Applied Linguistics* 14.2: 107-120.
- MARTEL, PIERRE. 1991. Conference opening. *Actes du colloque sur les anglicismes et leur traitement lexicographique: Communications, discussions et synthèses: Magog du 24 Au 27 Septembre 1991*, ed. by Pierre Martel and Hélène Cajolet-Laganière, 297-301. Quebec: Gouvernement du Quebec, Office de la langue française.

- MOUGEON, RAYMOND, and ÉDOUARD BENIAK. 1991. *Linguistic consequences of language contact and restriction: The case of French in Ontario, Canada*. Oxford: Clarendon Press.
- NADEAU, JEAN-BENOIT, and JULIE BARLOW. 2003. *Sixty million Frenchmen can't be wrong: Why we love France but not the French*. Naperville, IL: Sourcebooks, Inc.
- NADEAU, JEAN-BENOIT, and JULIE BARLOW. 2006. *The story of French*. New York: St. Martin's Press.
- PICONE, MICHAEL D. 1996. *Anglicisms, neologisms and dynamic French*. Amsterdam: John Benjamins Publishing.
- REY-DEBOVE, JOSETTE and GILBERTE GAGNON. 1980. *Dictionnaire des anglicismes: Les mots anglais et américains en français*, v-xvi. Paris: Robert.
- RIFELJ, CAROL. 1996. False friends or true?: Semantic anglicisms in France today. *The French Review* 69.3: 409-416.
- ROBERT, PAUL. 2005. *Le nouveau petit robert: Dictionnaire alphabétique et analogique de la langue française*. Paris: Dictionnaires Le Robert.
- SPENCE, NICOL C. W. 1989. Qu'est-ce qu'un anglicisme? *Revue de Linguistique Romane*, 53: 323-334.
- TANNEN, DEBORAH. 1982. Oral and literate strategies in spoken and written narratives. *Language*, 58.1: 1-21.
- THÉORET, MICHEL. 1991. La situation des anglicismes au Québec. *Actes du colloque sur les anglicismes et leur traitement lexicographique: Communications, discussions et synthèses: Magog du 24 Au 27 Septembre 1991*, ed. by Pierre Martel and Hélène Cajolet-Laganière, 79-92. Quebec: Gouvernement du Quebec, Office de la langue française.
- THOGMARTIN, CLYDE. 1984. Some "English" words in French. *French Review* 57.4: 447-455.
- TIMMINS, STEVE. 1995. *French fun: The real spoken language of Québec*. Toronto: Wiley.
- TRESCASES, PIERRE. 1982. *Le Français vingt ans après: Langue et société*. Montréal/Toronto: Guérin.
- ULLMANN, STEPHEN D. (1947). Anglicisms in French: Notes on their chronology, range, and reception. *Publications of the Modern Language Association of America* 62.4: 1153-1177.
- VILLERS, MARIE-ÉVA DE. 1988. *Multidictionnaire des difficultés de la langue française*. Montreal: Editions Quebec/Amerique.
- WEINREICH, URIEL. 1974. *Languages in contact: Findings and problems* (8th ed.). The Hague: Mouton.
- WISE, HILLARY. 1997. *The vocabulary of modern French: Origins, structure and function*. London: Routledge.