A cross-linguistic corpus-assisted discourse study of language ideologies in Canadian newspapers

Rachelle Freake
School of Languages, Linguistics and Film
Queen Mary, University of London

1. Introduction

This paper examines language ideologies in Canadian newspapers using the method of cross-linguistic corpus-assisted discourse studies (CADS) (Baker et al., 2008; Freake et al., 2011; Partington, 2004). More specifically, this paper has three angles. The first angle is theoretical: it is argued that the Canadian context provides a useful site for cross-linguistic CADS. The second angle is methodological: it is argued that cross-linguistic CADS is a useful method for studying the Canadian context. The third angle is topical: three examples are provided from 2009 data that suggest some of the ways in which cross-linguistic CADS shed light on language ideologies in Canadian newspapers. This paper will begin by outlining the historical context of language in Canada, in particular with respect to language in the media. Next, the data will be presented and an overview of the CADS methodology provided. Finally, three examples of findings on language ideologies will be presented and the paper will conclude with a summary of the three foci (theoretical, methodological, topical) and directions for future research.

2. Historical context

Language has great significance in Canada and has had such significance for many centuries. Ever since Europeans first settled in Canada, language has come to symbolize societal – if not national – difference. The first Europeans to establish permanent bases in Canada were the French, who continued to comprise an important component of the Canadian population even after the British gained control of the territory in the 18th century. Marginalized by legal, educational, and religious systems that differed from English-speaking Canada, French speakers in the province of Quebec became aware of their distinctiveness and activists of their rights in the mid-twentieth century. The resulting Quebec nationalist movement and its associated discourse took shape almost exclusively in French, placing it at odds with discourse in the vast majority of rest of Canada. The differences existed not only in terms of the medium of communication (because English has long since been the dominant language in Canada outside Quebec), but also in terms of the (different) ideologies of nation and language embedded in an English Canadian nationalist discourse. To a large degree, English Canadians had become accustomed to, at best, forgetting the French population and, at worst, attempting to stifle their proliferation outside Quebec. Indeed, many English Canadians came to form communities so disparate and unrelated to those of French Canadians that they too arguably compose a nation. This nation had its own language (English), territory (Canada – with or without Quebec, depending on perspective), culture (adapted from Britain and the USA), and history unique to itself. Their nationalist discourse, too, tended to be expressed only in English.
These two oppositional versions of nationalism (Quebec and English Canadian) continue to exist in Canada today, each with its corresponding nationalist discourse. The result is that both the Quebec national discourse and the English Canadian national discourse exist in singular languages and each discourse comprises unique understandings of the role that language does (or at least should) play in society. For example, in the Quebec nationalist discourse, the French language is highlighted as having a historical and pivotal role as an identity marker for the Quebecois, and this is taken as a topos (“conclusion rule”) for explaining why French must be defended against the infringements of all other languages (Bouchard, 2002; Heller, 1999; Oakes & Warren, 2007). In contrast, the English language is taken for granted in English Canadian society to the extent that it is either unremarked or serves a primarily functional (i.e. communicative) role in society, which is at odds with the integrative and emotional role that French plays in the Quebec nation (Boberg, 2010; Heller, 2003; Resnick, 1977, 1994) (these themes will be discussed in more detail in Section 5).

Thus, Canada is home to a divided people, and the division has two constituent parts. On the one hand, because most Canadians are monolingual in terms of the official languages (English and French), they are separated by the medium of communication: English speakers and French speakers speak different languages. On the other hand, Canadians are separated by disparate understandings – that is, ideologies – of the role that language does or should play in society, and these are tied to the two dominant versions of nationalism that exist in Canada. Canada is, then, is divided both linguistically and nationally: it is divided both according to the medium of communication and according to the national values and roles attributed to the medium of communication.

3. Theoretical background

This paper investigates the connection between language ideologies and nationalist discourses in Canadian newspapers. Here, language ideologies are understood to mean systematically-held, socially shared beliefs about the role that language does or should play in the nation (Boudreau and Dubois, 2007; Woolard, 1998). Nationalist discourses are taken to mean overarching, socially-indexing, semiotic constructions of the way nations are understood (i.e. ideologized) to be meaningful categories of belonging (Blackledge and Pavlenko, 2002; Blommaert, 1999; Wodak et al., 2009). Language ideologies are connected to nationalist discourses because, since the 18th century, languages have often been understood as markers of difference in the “one-language-one-nation” trope. In Canada, this language-nation complex is confounded because understandings of languages and their place in the nation differ according to the medium of communication. What’s more, language ideologies and national discourses are well-known to be itinerant in the news.

Previous research has indicated that the media serves as a useful site for the examination of national and linguistic ideologies. Media language can have a profound effect on the way populations speak, inform themselves, and even think (van Dijk, 1993), effectively creating “imagined communities” of belonging (Anderson, 1983; Billig, 1995). Indeed, research in Canada has suggested that the French and English Canadian media discursively isolate the population by cultivating cross-cultural misunderstandings (Oakes & Warren, 2007; Taras, 1993). The argument is that because the Quebec nationalist discourse exists primarily in French and the English Canadian nationalist discourse exists primarily in English, when these discourses are embedded in the media, the predominantly monolingual population is exposed to only one national discourse and the “other” is shrouded in mystery: English speakers and
French speakers in Canada are, as Hugh MacLennan (1945) so famously wrote, two solitudes.

Despite the widely-recognized role of the media and the oft-lamented solitudes, remarkably little research has endeavoured to account for differences between Canada’s French and English-language media. There have been surprisingly few comparative discourse analyses of French and English Canadian texts (for exceptions, see Gagnon, 2003; Kuhn and Lick, 2009; Robinson, 1998). Hence, this paper addresses this research gap by examining the ways in which Canada’s official languages (English and French) are represented and used to serve different ideological purposes in the nationalist discourses that are circulated in the English and French Canadian media (Fletcher, 1998). The specific research question being addressed, then, is the following: how are language ideologies and national discourses, as evident in Canadian newspapers, similar or dissimilar in English and French? This question entails a methodology that is able not only to capture the large-scale and small-scale differences between the two languages, but also able to systematically compare the large amounts of data required for an investigation that is as balanced, representative, objective, and thorough as possible. The methodology that was found to be most apt for this study is cross-linguistic corpus-assisted discourse studies.

4. Methods and data

Cross-linguistic corpus-assisted discourse studies (CADS) involves the application of CADS to multilingual data. In other words, it involves the application of the theory and methods of both corpus linguistics and discourse analysis to multilingual data. Previous research has demonstrated the strength of the CADS approach (see e.g. Baker et al., 2008; Morley & Bayley, 2009; Partington, 2010). Corpus linguistics, for one, enables researchers to uncover broad discursive patterns through frequency, statistical significance, and language patterning alignment techniques. Discourse analysis, on the other hand, involves the in-depth analysis of concordance lines, clusters and whole articles. The corpus methods used here draw on the work of Baker (e.g. 2004, 2006, 2010), Partington (e.g. 2004, 2009), Gabrielatos (2007) and Stubbs (e.g. 1996, 2001) and Mike Scott’s WordSmith Tools program is used for analysis. The discourse analysis used here is based on Hallidayan functional grammar (Halliday & Matthieson, 2004) and critical discourse analysis (Fairclough, 1989; Wodak et al., 2009; van Dijk, 1993, 2006). As discussed, CADS has been used by numerous researchers in recent years, and it has proven to be a valuable way to approach and analyze data. However, CADS has rarely been used on multilingual corpora (for exceptions, see Freake et al., 2011; McEnery and Salama, 2011; Qian, 2010). In this paper, the application of CADS on bilingual corpora was found to be useful, and sheds light on how language ideologies and nationalism differ according to the medium of communication in Canada.

The data used in this research include over 27 000 English and French articles (over 11 million words) from a three-week period in 2009 that are drawn from 17 newspapers (12 English and 5 French) with the highest circulation figures across the five regional divisions of Canada (see Canadian Newspaper Association, 2009). The breakdown of each corpus can be seen in Table 1.
Table 1. English and French corpus breakdown

The corpora were designed to be relatively large so that they would cover a time-period that was diverse enough so that topics and events would be diverse. In this way, there would be fewer singular topics or events that would dominate the content of the corpus. The aim was to have a broad time period so that a more general sample of news topics would be covered so that ideologies – in particular language ideologies – could surface with regard to these topics. The corpora were also designed to cover a time period (June 15-July 7) that included both of Canada’s dominant national holidays, one which is predominantly English Canadian (Canada Day, July 1), and one which is predominantly French Canadian (St. Jean Baptiste Day, June 24). Since national days are manifestations of nationalism (Hayday, 2010; McCrone & McPherson, 2009), and since Canadian nationalism is heavily imbued with language ideologies, this time period provides an ideal opportunity to study the interrelatedness of national discourses and language ideologies in Canadian newspapers.

One of the many challenges that emerged with the use of relatively large corpora in different languages was the inability to directly and systematically compare and contrast findings from each language. Cross-linguistic corpus-assisted discourse studies (CADS) was found to be a useful method for uncovering discourses and ideologies that tended to differ between Canada’s two dominant language groups.
5. Findings

In this section, three findings are highlighted that shed some light on how the cross-linguistic CADS approach revealed differences in language ideologies and nationalism between the English and French Canadian newspapers. The first section discusses findings on anglonormativity, the second section discusses findings on hyperlinguistic awareness, and the third section presents findings with regards to language endangerment.

5.1 Anglonormativity

One of the most salient findings that emerged first and foremost from the data indicated a strong sense of anglonormativity in the English corpus. Here, the term anglonormativity is used to refer to the naturalization and embeddedness of the English language. This finding became apparent first in the quantitative findings of the corpus, but was later substantiated by further qualitative findings. In the English corpus, the terms FRENCH and ENGLISH occur 1490 and 791 number of times, respectively. Notice that the term ENGLISH occurs around half as often as FRENCH, even though the medium of the corpus is English. In contrast, in the French corpus, the relationship between the French equivalents of these words (FRANÇAIS, ANGLAIS) is the inverse (N=1139, 368, respectively). In other words, FRANÇAIS occurs nearly three times as often as ANGLAIS. Also noteworthy is the fact that, because of the different sizes of the corpora (the English corpus is more than double the size of the French corpus in terms of words and articles), these frequencies indicate the different saliency of language in each corpus.

Discussions of languages are not limited to the use of the words “French” and “English”, however. There are a constellation of terms used to refer to language issues in Canada (see Tables 2 and 3).

| • ANGLO, ANGLOS, ANGLICIZE, ANGLOPHONE, ANGLOPHONES  | • ANGLAIS, ANGLAISE, ANGLAISES, ANGLICISME, ANGLICISE |
| • BILINGUAL, BILINGUALS, BILINGUALISM | • BILINGUE, BILINGUES, BILINGUISE |
| • ENGLISH | • FRANÇAIS, FRANÇAISE, FRANÇAISES |
| • FRANCO, FRANCOPHONE, FRANCOPHONES, FRANCOPHONIE | • FRANCO, FRANCOS, FRANCOPHONE, FRANCOPHONES, FRANCOPHONIE |
| • LANGUAGE, LANGUAGES | • LANGAGE, LANGAGES, LANGAGIER, LANGAGIÈRE, LANGAGIÈRES |
| • LINGUISTIC, LINGUISTICS | • LINGUISTIQUE, LINGUISTIQUES |
| • MONOLINGUAL, MULTILINGUAL, UNILINGUAL | • LANGUE, LANGUES |

Table 2: Language terminology (English)

Table 3: Language terminology (French)
If these constellations language terminology are taken into consideration, combined, the words are contained in 15.9% of articles in the French corpus, but in only 7.86% of articles in the English corpus. In other words, the French corpus contains a higher percentage of newspaper articles that refer to language than the English corpus.

In terms of more qualitative findings, discursive patterning contrasts French (FRENCH/FRANCOPHONE/S) in Quebec with “the rest of Canada”, which is implied (rather than overtly stated) to be English-speaking (see Table 4).

<table>
<thead>
<tr>
<th>Table 4: QUEBEC contrasted with THE REST OF CANADA</th>
</tr>
</thead>
<tbody>
<tr>
<td>In addition, other examples in the corpus highlight the pattern OUTSIDE + QUEBEC, where the majority of examples highlight the unusualness of the use of French outside of the French-dominant province (see Table 5).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5: FRENCH/FRANCOPHONE and OUTSIDE QUEBEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evidence of hyperlinguistic awareness first became evident in the frequencies of the French corpus, which contrasted with those in the English corpus. As discussed in Section 5.1, whereas in the English corpus the terms FRENCH and ENGLISH occur 1490 and 791 number of times, respectively, the equivalent terms FRANÇAIS and ANGLAIS occur in inverse proportions in the French corpus (N=1139, 368, respectively). In other words, the French language is highlighted significantly more than the English language in the French corpus. This is evident, again, when the constellations of language terminology (see Tables 2 and 3) are sought out in the corpora. The proportion of French articles that contain at least one reference to language terminology (15.9%) is twice that of the English articles (7.86%). There is also ample evidence of hyperlinguistic awareness in qualitative findings. For example, numerous references (43 occurrences) highlight the repeated collocation between</td>
</tr>
</tbody>
</table>

**5.2 Hyperlinguistic awareness**

The next finding is “hyperlinguistic awareness”, and this builds directly on findings related to anglonormativity. Hyperlinguistic awareness refers to the heightened and often marked awareness of linguistic issues, which in this case, tends to be centred around the French language. The hyperlinguistic awareness in the French corpus contrasts with the anglonormativity – that is, the embeddedness and commonsense nature of language issues – in the English corpus. The existence of hyperlinguistic awareness in French is hardly surprising given the history of French in Canada, including the fight for French language education rights and the Quebec nationalist movement (see e.g. Bouchard, 2002; Hayday, 2005; Oakes and Warren, 2007).
5.3 Language endangerment

The final finding that will be highlighted here is language endangerment, which here is taken to mean the concern over the status and future of a language. Discussions of language endangerment take place in both the English and French newspapers, where there is repeated and explicit mention of the French language being “threatened” (menacé, 5 occurrences; “threatened”, 8 occurrences). In addition, though, language endangerment is also discussed in more subtle ways – especially in the French corpus. There, references to the need to “defend” (défendre), “promote” (promouvoir) and make Quebec “more French” (plus français) permeate the corpus (see Table 8).
The English language, French-English bilingualism and immigrant languages are seen as threats. This is because English, as a powerful international language, has traditionally been the language adopted by immigrants to Quebec (see e.g. Anctil, 2007) (see Example 1).

Example 1

“It is this language [English] that constitutes a threat. Still today [...] neither Chinese, nor Portuguese, nor any other language spoken in Quebec, except English, threatens French…”

c’est cette langue [l’anglais] qui constitue une menace. Aujourd’hui encore [...] ni le chinois, ni le portugais, ni les autres langues parlées au Québec, sauf l’anglais, ne menacent le français…

(Le Devoir, June 22, 2009, p. a7)

In contrast, though, findings from the English corpus show that while language endangerment is often discussed, it tends to be discredited and mocked. This is achieved, for example, through the use of quotation strategies, modalization, interjections, and rhetorical questions (see Examples 2 and 3).

Example 2

A prime example of this has occurred as the sponsor of an “alternative” St-Jean-Baptiste Day celebration has decided that two English-language acts, scheduled to appear, may not do so because their presence might confuse people and pose a “threat” to the French language in Quebec.

McDevitt, M., “For all the world to see”
(The Record, June 16, 2009, p. 6)

Example 3

Did your article really say a bluegrass group and a country singer were banned from the St. Jean Baptiste Day celebrations because their singing in English would constitute a threat to the French language? For crying out loud, the lyrics of bluegrass and country music are a threat to the English language. For that reason alone, they are wildly popular to English speakers.

Anonymous, “Threat to whom?”
(The Gazette, June 17, 2009, p. a20)

Results show, then, that French and English Canadian newspaper project different perspectives on language endangerment. This is perhaps unsurprising, since the blame for French language endangerment is nearly always placed on the predominance of the English language, as Example 1 clearly demonstrated. These examples all taken together indicate that French language endangerment is not taken seriously in English newspapers, and while some
times endangerment is subtly discredited, other times it is openly mocked. In contrast, in the French corpus there appears to be a consensus that the French language is endangered or at least in need of protection from English.

6. Conclusion

In conclusion, then, this paper has discussed language ideologies in Canadian newspapers. The paper aimed to study how language ideologies were similar or dissimilar in English and French Canadian newspapers. Results showed that while anglonormativity dominates the English newspapers, hyperlinguistic awareness dominates the French newspapers. Although both English and French Canadian newspapers discuss language endangerment, the French newspapers present the English language, bilingualism, and even foreign languages as threats to the French language, which needs to be protected. The English newspapers, however, tend to mock or attempt to discredit language endangerment.

This paper has also aimed to provide three angles to the topic. With regards to the theoretical angle, the Canadian context has proven to be a useful site for using cross-linguistic CADS because it is an authentic site where different discourses, each in a different language, exist within a single context. The Canadian situation is a valuable one to study using cross-linguistic CADS, then, because it allows a researcher to compare discourses in different languages that emerge from ostensibly similar contexts (i.e., comparable media producers within a single country). With regards to the methodological angle, this paper has presented findings that emerged because of the strength of the cross-linguistic CADS approach. This approach allowed large corpora in different languages to be mined for comparisons not only in terms of frequency and statistical significance, but also in terms of discursive constructions and representations. Finally, with regards to the topical angle, this paper has shown that there are important differences between language ideologies embedded in the English and French media. These ideologies arguably pertain to some of the differences that exist between English speakers and French speakers in Canada, and the ways that these groups of people relate to the country as a whole. Although Canada is officially bilingual, if the media language reflects language ideologies that are held more widely in Canadian society, it may be that (predominantly monolingual) speakers of each language could have different understandings of the role that languages play in the country. In other words, this study of language ideologies in the Canadian press could have more wider implications on the significance of Canada’s national and linguistic divide and the ways in which individuals conceptualize their belonging in the country. This, of course, must be investigated in language samples that go above and beyond media language.
7. References


