Subordinate clause marking in Montreal Anglophone French and English

Hélène Blondeau

University of Florida

210 Dauer Hall
P.O. Box 117405
Gainesville, FL 32611-7405

Blondeau@rll.ufl.edu

Naomi Nagy

Department of Linguistics
University of Toronto
130 St. George Street, Room 6076
Toronto, Ontario M5S 3H1

nn.in.nh@gmail.com

Hélène Blondeau, University of Florida
Naomi Nagy, University of Toronto[1]

1. Introduction

In 1993, Gillian Sankoff and Pierrette Thibault began an innovative investigation of variation in the speech of young Anglophone Montrealers (AM), with an aim toward understanding how the degree and types of exposure to French influenced the way these speakers use French. This ambitious project benefitted from the years of experience with sociolinguistic investigation in multilingual communites, and in Montreal in particular, that Gillian brought to the project, and we are very pleased to have had the opportunity to work with her on this project and to now have this opportunity to thank her for both her earlier groundbreaking research and her excellent skills as a mentor. We are grateful to her for conceiving of this important project and for allowing us access to the data, but even more for having set the stage in so many ways for the study of the interaction of linguistic and social variation and for having trained so many of us who work in this area, including Pierrette Thibault, the co-PI for the original Montreal Anglophone project, and both authors of this paper. Since this study began, many aspects of both the French and English of these speakers have been investigated. An overview may be found in Blondeau et al. (2002). In this chapter, we aim to follow Gillian's example of examining linguistic variation in its social context, as we investigate one more pattern of variation in AM speech.

French and English both permit the variable presence of overt complementizers as the head of subordinate clauses. English allows the presence or absence of the complementizer *that* as illustrated in (1) and French allows the presence or absence of the complementizer *que* as illustrated in (2). Previous studies show that this variable is constrained by a set of linguistic and extra-linguistic constraints in each language (cf. Biber 1999 for English, and for a recent account of a variety of Quebec English see Torres Cacoullos & Walker *forthcoming*; for a variety of Quebec French, see Dion 2003). Since French is now an integral part of the linguistic repertoire of Anglophone Montrealers, we compare the patterns of usage of our bilingual speakers to both of these other studies, looking at similarities and differences in the effects of linguistic

constraints. Specifically, we analyze and compare the linguistic factors that contribute to the variable presence of overt complementizers (COMP) in the English L1 and the French L2 included in the linguistic repertoire of young Anglophone Montrealers, who produce tokens such as those shown in (3) and (4).

1. L1 English examples (from the *Quebec English Corpus*, Poplack, Walker and Malcolmson 2006, cited in Torres Cacoullos & Walker *forthcoming*: 1)

    a. And I let it slip that Darth Vader was Luke's father.
    b. I can't even believe Ø I just said that.

2. L1 French examples (from the Ottawa-Hull Corpus (Vieux-Hull neighborhood), Poplack 1989, cited in Dion 2003: 12).

    a. *Moi je leur disais Ø c'est pareil comme nous autres si on irait dans leur pays.*
        I told them Ø it's the same, like we—if we went in their country.
    b. *Bien parce qu'elle disait que les filles étaient pas assez distinguées.*
        Well, because she said that the girls weren't distinguished enough.

3. Anglo-Montrealer example in French

    Je pense Ø c'est plus anglophone je pense que les compagnies sont: donnent un sens de anglophone. (Greg)
    I think Ø it's more Anglophone. I think that the companies give a sense of being Anglophone.

4. Anglo-Montrealer example in English

    I think Ø he thought Ø it was really cool that I spoke French and that I was bilingual.
    (Liz)

In current spoken English it is common to have subordinate clauses introduced without an overt COMP, as in (1 and 4). The absence of COMP, although possible, is less common in French (as in 2 and 3). These AM speakers respect this difference: they exhibit only 27% COMP presence in Englishbut 77% in French. Their rate of usage in English is higher than the ~15% reported by Biber (1999: 145) for the conversational register, but given methodological differences, the difference is not noteworthy. In our analysis of the varation, the factors examined include the form and reference of the matrix and subordinate clause subjects, identity of the subordinating verb, frequency of collocation of these two as well as lexical frequency of the matrix verb, phonological context, and several measures of semantic and syntactic complexity. Throughout, we present the application value as COMP presence. That is, we compare how often there is a surface *that* or *qu*e. This makes our study compatible with other studies of the variable in English, although French studies tend to report the rate of COMP absence.

After a literature review summarizing the constraints on variation in the two language varieties, we explicitly circumscribe the variable context for our study and define the linguistic variables to be examined. The next section of our paper provides an analysis of the similarities and differences both within the speech of these Anglophone Montrealers and with other recent studies of the same variable. Then we briefly explore how *like* could be reanalysed as a COMP since it appears in the same surface position as the (deletable) COMP, and make predictions for the evolution of *comme* in French. This provides an analysis for sentences like:

5. Because I felt like, Uh: OK, I understand that the French people have to protect themselves. (Louisa)

We then position our analysis within the social context of the variation.

2. Previous studies of the variable presence of COMP

2.1. French

The presence or absence of the COMP in subordinate clauses has been examined from a number of variationist perspectives. In her seminal article, "A quantitative paradigm for the study of communicative competence," Sankoff refers to the first analysis of the deletion of COMP *que* in Montreal French (Sankoff et al. 1971, cited in Sankoff 1980: 66). She uses this example to emphasize the structure of variable phonological deletion constraints. In another article devoted to the role of phonology in variable rules, Sankoff and Brown (1976) refer to another construction containing the form *que* in co-occurrence with wh-forms.[2] *Que*-presence is variable not only for Quebec French but also for other varieties of French, as commented upon in Gadet (2007), making *que* an interesting candidate to examine from a variationist perspective. In this article we restrict ourselves to the variable presence of the COMP *que*.

Most researchers agree that the variable presence of COMP *que* in French is constrained by social factors and that the absence of *que* is stigmatized. However for many years there has been debate over the role of linguistic factors in the explanation of the variability. Three types of conditioning are identified: phonological, syntactic and lexical, but the relative importance of each is disputed.

Sankoff et al. (1971) and Sankoff (1980) first looked at this variable, identifying the phonological environment as the major constraint on the variation. Their analysis of the distribution of the variable among sixteen Montrealers led them to identify the preceding and following phonological environment as relevant factors. The analysis identified sibilants as a favorable context for *que* deletion, and the interpretation provided was one of consonant cluster simplification (Sankoff 1980: 66). Working with the same data, Connors (1975) proposed an alternative explanation involving syntactic factors. According to her, the deletion was conditioned by the nature of the subject in the subordinate clause. If a pronoun appeared in the subordinate clause rather than a noun, a null COMP was favored. According to Connors, the phonological environment was an epiphenomenon of the syntactic effect, since several frequent subject pronouns in French start with a sibilant (i.e. *je* 'I' and *ça* and *ce* 'that'). The debate remained open until new studies were undertaken.

Ten years later, a study of the same variable in the French of Ottawa-Hull was undertaken by Martineau (1985), followed by a real time study of Montreal French (Warren 1994). Both found that the less sonorant the following segment was the more the deletion of *que* was probable, at least partially supporting Sankoff's interpretation rather than Connors.' However they also both discarded the effect of the phonological environment preceding the COMP and indicated a possible lexical effect. Warren also pointed out the importance of a lexical effect and discussed extensively the specificity of COMP deletion following the collocation *je pense*.

In a recent real-time study on the same variable in Canadian French, Dion (2003) tried to reconcile differences in reported signficant factors by carefully examining the interaction between factor groups. She was able to convincingly disentangle the effect of each factor and identified only two factors contributing independent effects. In the preceding context, the lexical identity of the matrix verb plays an important role in determining the likelihood of deletion. In the following context, Dion discarded the syntactic explanation in favor of a phonological one, since all pronouns did not behave in the same way. Pronouns starting with a vowel clearly disfavor deletion, while pronouns starting with an initial consonant favor it. The NP's were at an intermediate point but this was also related to their phonological status, most of them starting with nasals and liquids. Her demonstration therefore convincingly supports Sankoff's, Martineau's and Warren's accounts, and her findings were further supported by additional data presented in King & Nadasdi's (2006) report on Acadian French *que*.

2.2. English

The variable presence of complementizers in English has been well documented in a number of publications. The most recent that we are aware of is Torres Cacoullos & Walker (*forthcoming*) and we refer readers to Section 2 of that paper for a thorough review of earlier work as well as a succinct history of the development of the COMP *that*. Torres Cacoullos & Walker present their study of this variable in the Quebec English Corpus (Poplack, Walker & Malcolmson 2006) in a multivariate framework similar to ours (cf. Nagy & Blondeau 2005) but are innovative in their investigation of the possibility that certain collocations that frequently occur between the subject and verb of the matrix clause may in fact have grammaticized into discourse markers, at least in

some of their occurences. They make the distinction that a sentence can truly be conceived of as containing a matrix and subordinate clause only when there is semantic content on both sides of the COMP. When there is no content in the putative matrix clause, the collocation may be better analyzed as a discourse marker, with typical examples being *you know* and *I think*. Their quantitative analyses of ~3,000 tokens, considering common and rarer collocations separately, convincingly support their arguments. This is relevant to our analysis as they found that the COMP is much less likely to surface when the "matrix clause" consisted of these common collocations, although the same linguistic factors constrain the variation in COMP presence in both common and rarer structures.

3. Methods

In this section, we briefly summarize the methods used to collect our data. They are presented more completely in Sankoff et al. (1997) and Blondeau et al. (2002). We then turn to the specific manner of coding and analysis to show how COMP behaves in our corpus and how this compares to its behavior in the studies mentioned in the preceding section.

3.1. The speakers[3]

The speakers are 18 Anglophones born in metropolitan Montreal. Their ages were 20-34 when they were first interviewed (1993-1995). Many of the speakers are a self-selected group of volunteers. The remainder are a sample of the 1990 graduating class of an English language high school on the border between an Anglophone and a bilingual neighborhood. The school offers two levels of French immersion as well as French taught as a subject (for those less advanced in French). It had recently received an influx of Francophone speakers. All speakers lived in the Greater Montreal area at the time of the interview and spoke English with their parents. They differ, however, in their mode of acquisition of French, the type of exposure they had to French as children, their current degree of contact with Francophones socially and in the workplace, and in the degree to which they use French in their daily lives.

3.2. Corpus development

The self-selected participants were recruited by newspaper ads seeking bilingual speakers placed in two free Montreal newspapers. The high school graduates were located by cross-referencing their high school yearbook and a telephone directory. Potential participants were telephoned and asked (in English) if they would participate in an interview about bilingualism in Montreal.

Each participant was interviewed first in French by a Francophone and a few weeks later in English by a non-local Anglophone. Topics for both interviews included school and family background, use of French in various contexts, attitudes toward French politics, people, and culture, and incidents where language differences played a significant role. Each interview lasted about one hour. All interviews were tape-recorded and fully transcribed. All our data is, therefore, from the conversational register, as defined in Biber (1999).

3.3. Methods of analysis

We first examined the frequency of subordinate clauses in the corpus. We then conducted a variety of multivariate logistic regression analyses using Goldvarb X (Sankoff et al. 2005) to determine the relative effects of the linguistic factors on the rate of COMP presence in various subsets of our data. Once a set of significant linguistic factors was determined, social factors were added in. Separate analyses were conducted for the French and English data. Goldvarb X produces factor weights indicating the strength of each factor, or, in essence, how likely a particular sentence is to contain a surface COMP if it has a particular linguistic or social attribute, independent of its other attributes. Factor weights were calculated using the "one-level" analysis, and the relative significance of each group was determined using the "step-up, step-down" analysis and consideration of the range of weights of the factors.

3.4. Linguistic data coding

All sentences illustrating the variable presence of a COMP as the head of a subordinate clause were extracted from the corpus, providing about 500 tokens in French and about 1,000 in English. We also extracted all examples of English *be like* (N ≈ 90) and French *être comme* (N ≈ 20) used to introduce reported speech, for reasons explained in Section 6. Sentences such as those shown in (6) were excluded, as variation in COMP presence is not possible.

6.   a.   I never know like [*that/Ø which one is which] (Liz)

       b.   It's [that/*Ø the person speaks both and I speak both] (Louisa)

       c.   And I love [that/*Ø I can. Um:] and [that I can read, just at all, y'know, just that I'm literate in two languages.] (Louisa)

       d.   *Ça dépend [qu'/*Ø est ce qui joue]* (Joan)

          That depends who's playing.

Each subordinate clause was coded for the presence or absence of a COMP and for the independent internal and external factors described below.

3.5. *Dependent variable*

Each token sentence was first coded for the dependent variable: presence or absence of a COMP at the head of the subordinate clause. In English, *that* and *like* were considered as surface COMP forms, but the *like* tokens are relegated to a separate analysis, discussed in Section 6. In French, the surface COMP forms are *que* and *qu'*, and *comme* (again, see Section 6).

3.6. *Independent linguistic variables*

The list of independent variables is provided in the appendices, and each possible variant is exemplified by a sample sentence from our corpus. The distribution of forms for each variable is shown in Tables 2 and 3, discussed below. The variables are grouped into four categories: those related to lexical identity and frequency, syntax, phonology, and semantics. Several earlier studies have tried to tease apart the effects of lexical identity of the matrix clause verb and its frequency. To contribute further data to this complex question with important ramifications, we

coded for both. The syntactic factors relate to the structural complexity of the sentence and are included to bolster previous reports that COMP is more likely to surface in more complex sentence structures (Torres Cacoullos & Walker *forthcoming*: 11, Biber 1999). Semantic factors are similarly important, in that previous reports suggest that COMP will surface more often when there is less semantic cohesion between the matrix and subordinate clause, that is, less co-referentiality of the two parts (*ibid*). Phonological factors are included because of the possibility that COMP presence/absence is (partially) determined by marked vs. unmarked syllable structure, as has been shown for French.

3.7. *Independent social variables*

We have selected speakers who differ in terms of their means of acquisition of French, degree of contact with French, and ability with prescriptive French rules. For ease of comparison, five types of information are shown for each speaker in Table 1. The "Contact" column indicates the degree to which they have been exposed to French. "High" indicates the most contact, whether it be through school, other activities, or friends and family; and "low" indicates least. "Education" focuses on the type of schooling they received. 1 point was granted for attending an English school with French as a subject. 2 points were granted for partial attendance in a French medium or immersion school. 3 points were granted for full attendance in a French medium school or (post-) immersion program. Students who changed program type receive intermediary scores. Scores of 5 and above are shown as "high;" scores below 3 are "low;" others are "medium." The "Grammar score" indicates the percentage of nouns (out of 20) that they correctly and unambiguously marked for gender in a stretch of their French interview, included to show their general level of proficiency. Aside from the Grammar score, information is based on self-reports.

@@Insert Table 1 here.

4. Data

In this section, we look first at the distribution of subordinate clauses across the two languages and across speakers. We then define the envelope of variation with respect to the different linguistic variables considered. Finally, we turn to an examination of the frequency of several elements within the token sentences.

4.1. *Overall frequency of subordinate clauses*

Our first observation is that the overall frequency of subordinate clauses does not differ greatly between the English- and French-language interviews in our corpus. Thus, the cross-linguistic differences discussed below may not be discounted by claiming that the speakers' language is less syntactically complex (in the relevant manner) in their L2 than their L1. There are ~4.5 subordinate clauses per 1000 words in English, and ~3.5 in French. In the English interviews, this value ranges from 1.2 (Tony) to 9.7 (Don), and the range does not appear to correspond to any particular social features. The rate of COMP presence in the two languages does not appear to be correlated for individual speakers, and is always greater for AMF than AME, as Figure 1 illustrates.

@@Insert Figure 1 here.

The English data is comprised of 1,013 sentences containing a subordinate clause. The distribution of tokens according to all the factors considered is shown in Tables 2 and 3.

@@Insert Table 2 here. [If too long for 1 page when typeset, break into separate tables for (2a) *Verb identity and frequency factors; (2b) Syntactic factors; (2c) Semantic factors; and (2d) Phonological factor.* OR repeat headings on top of each page.]
@@Insert Table 3 here. [If too long for 1 page when typeset, break into separate tables for (3a) *Verb identity and frequency factors; (3b) Syntactic factors; (3c) Semantic factor; and (3d) Phonological factors.* OR repeat headings on top of each page.]

4.2. *Frequency of matrix verbs*

The lexical frequency of the different matrix verbs may be relevant in accounting for the variable patterns, at least in English (cf. Berkenfield 2001, Torres Cacoullos and Walker *forthcoming*). Tables 4 and 5 provide separate counts for common collocations vs. other uses of frequent verbs. In addition to calculating the relative frequency of each of the matrix verbs in our own corpus, we looked up the frequency of the verbs in the British National Corpus (Leech et al. 2001) and Corpaix (Véronis 2000), corpora of spoken language for which frequency statistics are available online.[4]

@@Insert Table 4 here.

@@Insert Table 5 here.

There are some noteworthy similarities in matrix verb usage between the two languages. Table 6 compares the most frequent matrix verbs in our corpora of AMF and AME. The same verbs are most common in both languages (*think, say, know* in English and their synonyms *penser, croire, dire, savoir* in French). This is true whether the frequent collocations of matrix subject *I* + verb are included (the additional values shown in parentheses) or not. Next most frequent, however, we see five verbs which are commonly used in one language, but not the other (*trouver* 'to find' *falloir* 'to be necessary', *voir* 'to see', and *sembler* 'to seem' in French; *tell* and *realize* in English).

@@Insert Table 6 here.

5. Multivariate analysis

Thus far, we have only presented raw numbers and percentages. We turn next to analyses that allow us to clarify the independent effects of each variable.

5.1. *Effects of linguistic factors in English*

Among the factors mentioned, there are a number which interact, so a binomial regression analysis cannot be conducted with all factors at once. First, animacy is fairly redundant with subject type (first and second person pronouns always refer to humans, third person masculine and feminine also refer to humans exclusively in the AME corpus and quite frequently in AMF). Inanimates only appear in the pleonastic/neuter or the NP categories. In a recoding of the AME data, in which animacy is only examined within these categories, it was found not to have a significant effect and is therefore not further considered.

Second, there are four factor groups which logically interact with each other. All of them relate to the matrix verb of the sentence (verb identity, semantic class, finiteness, and lexical frequency, as determined by the BNC). Four binomial analyses were conducted with just one of these factors each, and their results are compared in Table 7. (Just the weights for the one factor which was switched are presented.) As in all Goldvarb analyses presented in this paper, the application value is presence of the COMP.

@@Insert Table 7 here.

Of the four interacting factors that describe the matrix verb, the one which best fits the data is lexical frequency (having the smallest log likelihood). Therefore, we proceed with further analysis including this factor group and excluding the other three.

Table 8 provides factor weights for all significant factor groups in an analysis in which lexical frequency, is included, but the other three factor groups relating to matrix verb identity are excluded. Factors are listed in decreasing order of significance.

@@Insert Table 8 here.

5.2. *Effects of linguistic factors in French*

We next present the effects of these factors in French, and then integrate our discussion of the effects in both languages. The French data is comprised of 506 tokens. The rate of COMP presence is 77%. TA multivariate analysis was performed on the 506 tokens extracted from the French interviews. The multivariate analysis tested for the following five linguistic factors: preceding phonological environment, following phonological environment, subject of the matrix clause, subject of the subordinate clause and the lexical identity of the matrix verb. The results for the three factors selected as significant are illustrated in Table 9.

@@Insert Table 9 here.

The most significant factor is the lexical identity of the matrix verb, with a range of 58 points. The analysis identified five verbs disfavoring the presence of COMP: *sembler* 'to seem', *souvenir* 'to remember', *dire* 'to say', *penser* 'to think' and *croire* 'to believe'. Other verbs, such as *savoir* 'to know', *falloir* 'to be necessary', *voir* 'to see', *trouver* 'to find' and *être* 'to be', seem more favorable to the presence of COMP *que*, as do other infrequent verbs in our data.

There is significantly more COMP presence if there is a vowel or a liquid following the COMP site. Sibilant and stops instead disfavor the presence of COMP *que*. However, the preceding phonological environment was rejected from the multivariate analysis. The AMF speakers are not influenced (directly) by this factor in the choice of COMP presence.

As far as the syntactic factors are concerned, the subject of the matrix clause is significant (but with a lower range of 27 points). NP subjects are the most favorable to the presence of COMP *que,* followed by third person pronouns. In contrast, first and second person pronouns disfavor the presence of COMP. We checked if this was related to the collocation effect. For example, the verb *penser* 'to think' appears most of the time in cooccurrence with the pronoun *je* 'I'. This collocation represents 21% of the tokens (see Table 4) and is associated with 71% COMP presence. The second most frequent collocation, *je trouve* is associated with 74% COMP presence. For those two collocations, the rate of COMP presence is near the overall rate and therefore does not sufficiently explain of why the syntactic factor is selected by the analysis. Finally, the subject of the subordinate clause was not significant in AMF. We next compare multivariate analyses of the two languages.

5.3. *Anglophone Montreal English (AME) vs. French (AMF)*

The rate of COMP presence in the two languages differs. In AME, it reaches only 27%, while in AMF it is 77%. This difference of 50 percentage points confirms that the variable behavior of the informants regarding COMP usage is quite different in French and English. Comparing the general tendency with the most recent study available for L1 Quebec French (speakers living in Vieux-Hull, Quebec) we find a better match. COMP presence there is 63% for the younger speakers and 68% for the older speakers (Dion 2003: 28). Taking the younger speakers of Dion's study as the counterparts of our young Anglo-Montrealers, we observe a difference of 14 percentage points.[5] Therefore, from the strict perspective of relative rates of usage, one cannot see evidence of convergence or intersystemic influence. In other words, the lower rate of presence of *that* in English does not seem to interfere with the presence of *que* in French. In addition, we observe that the even higher rate of COMP presence in AMF in comparison with L1 French seems to reflect more conformity with the prescriptive norm of French, where *que* deletion is highly stigmatized. Comparing the input values for the two samples (see Table 10), we see again that there is a much higher rate of COMP presence in AMF than AME – it is not a spurious effect of the differential usage of different types of sentence constructions, etc., but a true difference between the grammars.

@@Insert Table 10 here.

The next section of Table 10 compares the factor weights for the various matrix verbs, in decreasing order of frequency of COMP presence in AMF. The nine verbs which appeared most frequently in the AMF corpus are listed individually, then the four next most common are grouped, followed by a grouping of the next 10 most common. Where the same verbs appeared frequently in AME, factor weights are listed. This table illustrates that not all of the same verbs were in the most common set in both languages and that their order in terms of preference for a surface COMP differs as well.

Turning to examine the effect of the following phonological environment, we see a difference between the two languages: less sonorant following segments favor the presence of the COMP in AMF, while more sonorant segments favor it in AME, when just a binary [+/-sonorant] distinction is made. As shown in Table 10, however, there is no consistent effect of sonority in the AME data, when finer-grained distinctions are made.

There is a different effect in the two languages for the subject of the matrix clause: NP's most favor COMP presence in AMF, while their weight is between that of first and second person vs. third person pronouns in AME.

In sum, it is not possible to say that the same grammar is used in the two languages – these speakers have acquired different grammars for AMF than for AME

5.4. *AMF vs. L1 French*

We turn next to a comparison of factor effects in L1 French to see whether these L2 speakers have in fact acquired the same patterns as native speakers. For both AMF and L1 French, the lexical identity of the matrix verb is an important factor. In Dion's (2003) findings, *sembler* and *penser* were also among the verbs that disfavor *que* presence. Our results also resemble Martineau's (1985) and Warren's (1994) findings which reported an association with the lexical identity of the verb. In particular, Warren (1994: 45), who devoted a qualitative analysis to frequent collocations with the verbs *dire* and *penser,* indicated that *je pense* was frequently associated with the absence of *que* in Montreal French. However, while *dire* strongly disfavors COMP presence in AMF, this verb was less associated with deletion in Dion's study of L1 French.

Noteworthy is the fact that the verb *falloir* favors *que* in AMF, but not in L1 French (Dion 2003). In our data the collocation *il faut* is the third most frequent (see Table 4). Dion found that *falloir* was very often used without *que* as illustrated in (7a). In our data, *falloir* instead favors *que,* as exemplified in (7b).

7.   a.   *Fallait Ø je me place en quelque part* (cited in Dion 2003: 9)
            It was necessary Ø I put myself somewhere.

b. *Il faut que je le fasse tu sais.* (Louisa)

    It's necessary that I do it, you know.

This might be related to the distinct syntactic nature of the English equivalent "have to + infinitive", a construction not involving the presence of a COMP. The constraints associated with this verb might not be acquired in the same way as other verbs with an equivalent structure in both languages. In addition, the strong correlation of *falloir* with the use of the subjunctive (Poplack 2001) may play a role. In the context of formal acquisition of French L2, the complete collocation *il faut que* is learned in association with the French subjunctive.

Similar results are reported for *falloir* in Acadian French where rates of *que* absence fall below average for this frequent verb (King and Nadasdi 2006). This verb is also among the most frequent verbs (Véronis 2000). However this verb is not associated with deletion in our data. At the other end, the verb *trouver* 'to find', which has medium frequency, is more favorable to deletion than the frequent verb *falloir*. Such observations do not support the association between verb frequency and deletion for the French data, as noted by King and Nadasdi (2006).

The subject of the matrix clause plays a role in AMF but with a weaker effect. While it is true that first and second person pronouns disfavor the presence of *que,* we hypothesize that this is related to the frequency of certain types of collocations, e.g. the verb *penser* 'to think' which is the most often used in collocation with the pronoun '*je'* in our data. This was reflected in Warren's study (1994), the only study that examines collocation effects in French. She devoted a qualitative analysis to three frequent expressions: *je pense* 'I think', *disons* 'let's say' and *parce que* 'because'.[6] In her data, *je pense* also favors the absence of *que*, paralleling our results. In our data the collocation *je trouve* was also associated with the absence of *que*, parralleling Warren's findings. Based on Thompson & Mulac's (1991) proposal, Warren considered these two constructions to be "unitary epistemic phrases." In her data set, these collocations are both associated with a higher degree of autonomy: they may appear in different positions in the utterence (Warren 1994: 47). In contrast, *je trouve* was not specifically associated with COMP absence in Dion's (2003) study.

Regarding the effect of the following phonological environment, AMF speakers exhibit parallel tendencies to L1 French speakers (Sankoff 1980, Martineau 1985, Warren 1994, and Dion 2003). There is no effect of the following syntactic environment in our data. This result may be explained by Dion's (2003) finding that the following syntactic environment factor interacts with the following phonological environment. Because the pronouns are not a homogeneous category, phonologically speaking, she considered the syntactic effect an epiphenomenon of the phonological one.

In sum, many of the linguistic constraints for AMF mirror those found significant in studies of French L1, confirming that these AM speakers have acquired the grammar of L1 Francophones.

5.5. *AME vs. Quebec City English*

Finally, we compare the AME grammar to that of another group of native speakers of English. This group lives in Quebec City and all qualify themselves as bilingual, "having learned French informally" in a city in which Anglophones have minority status (Torres Cacoullos & Walker *forthcoming*: 8). A majority of the speakers also report using French on a daily basis (Poplack, Walker & Malcolmson 2006: 199-201). They do not differ greatly from the AME speakers in terms of reported contact with French.

The first thing to note is the similarity in overall rate of COMP presence: 21% in Quebec City English (QCE) and 26% in AME. In both, the effect of verb identity is strong and significant, and similar verbs favor COMP presence in both languages. A second shared feature is the favoring effect of the presence of other morphemes between the matrix verb and the subordinate clause, a factor to which we return in Section 6. However, there are numerous differences between the Englishes of these two groups.

Our results differ from Torres Cacoullos and Walker's for the effect of both the matrix clause subject and the subordinate clause subject. While they report matrix clause noun subjects to favor *that* more than pronouns (Torres Cacoullos & Walker *forthcoming*: 21), we find a difference in the behavior of singular and plural nouns, and cannot combine them. Plural nouns

most strongly favor COMP presence, while singular nouns are the least favoring. Pronouns fall between. We also see salient differences across types of pronouns in AME, which they do not report: third person pronouns have a stronger preference for COMP than first & second person. As far as subordinate clause subjects, there is a significant effect: Nouns favor COMP most, *it* and *there* favor it least, with other pronouns falling in between) in QC (*ibid*: 21). Table 8 illustrates a different pattern for AME.

Two other factors emerge as significant in AME but not in QCE: the sonority of the phonological segment following the COMP (apparently not tested in QC) and polarity differences between the matrix and subordinate clause. (But, see Table 10 regarding the inconsistent sonority effect.) On the other hand, matrix clause finiteness and subordinate clause transitivity are not significant in the AME data but are in QCE (*ibid*.).

Table 11 presents a summary of the comparisons we have made. Shading highlights relevant similarities between speech varieties. Looking at AMF, the second language of our speakers, we can see that in several respects (overall frequency of COMP, the effects of the matrix verb, and the two phonological factors) it patterns like that of native speakers' French. There is no evidence that the AME grammar is interfering – no shared patterns between AME and AMF that are not also shared by L1 French. Interestingly, AME also does not pattern like QCE, except with respect to the syntactic complexity factor, which is shared across the two English varieties but not the two French varieties, the effect of the lexical identity of the verb, and the overall rate of COMP presence.

@@Insert Table 11 here.

5.6. *The problem with sentences like "I think ..."*

One concern that we had with our data is that certain collocations such as *I think* may not actually form a matrix clause that subcategorizes for a subordinate clause, but rather be a discourse marker or filler (much as *y'know* has come to be). In that case, the next clause is independent and could not be introduced by a surface COMP. It would, therefore, need to be excluded from the envelope of variation. In order to address this, we separated the very frequent forms from our corpus and conducted an analysis using only less common forms, which are not suspected of being on the lexicalization path toward discourse filler status. Following the method described by Torres Cacoullos and Walker (*forthcoming*), we conducted separate analyses of common [matrix subject + verb] collocations and rarer verbs to see if the same set of linguistic factors are operational in both. This allows us to determine whether, like Torres Cacoullos and Walker (*forthcoming*: 1), our data show that:

> [g]rammatical conditioning persists in fixed discourse formulas. Despite their high frequency and formulaic status, such formulas are not completely autonomous from the productive constructions from which they emerge... despite slight rates [of COMP presence], the linguistic conditioning of *that* in frequent collocations that behave like discourse formulas parallels its conditioning in the general construction.

This also allows us to contrast the patterning of COMP presence in AME and QCE in one other way. Table 12 shows the similarity of overall effect between AME and QCE. The input values for COMP presence in more vs. less common collocations are similar, and the differences between the two are nearly identical.

@@Insert Table 12 here.

However, when comparing factor effects in these two subsamples, we see different patterns in AME and QCE. The *only* significant independent factor in the AME analysis of common collocations is the presence of material intervening between the matrix verb and the subordinate clause (See Table 13, and see Table 5 for a list of the frequent collocations included). In the AME analysis that excludes the common collocations, the three significant factors are: lexical frequency, subordinate clause subject, and presence of intervening material (See Table 14). In each table, factor groups are listed in decreasing order of significance and only significant groups are listed.

@@Insert Table 13 here.

@@Insert Table 14 here.

In QCE, the same grammatical effects are evident in the common and rarer collocations. Torres Cacoullos and Walker (*forthcoming*: 28) report lexical identity, matrix and subordinate clause subject as significant in their subsample of common collocations. These are also the three most significant factors in their subsample that excludes the common collocations. Our data pattern differently from the Quebec City data. We do *not* see the same factors operating on the common and the rarer types of matrix clause verbs in AME. This is the most striking difference between the Montreal and Quebec City data. It may suggest that the AME variety has moved farther along toward grammaticization of the frequent collocations – and that the frequent collocations have in fact lost all trace of grammatical conditioning.

We are not in a position to definitely say that the grammars of QCE and AME are different, although we must leave that open as a possibility. Several other possible explanations exist for these different patterns. First, the QCE study included 2,820 tokens (*ibid*: 16) -- three times as many as our analysis. Our smaller data set also contributes to a methodological problem. Since *all* of the common collocations are of exactly the form "*I* + matrix verb," several factors must be omitted from the analysis. These include matrix clause subject type (always a first person pronoun), animacy of matrix subject (always [+human]), and matrix clause morphology (always finite). Additionally, among the common collocation tokens, there is not enough variety to be able to examine the effects of polarity (two variants must be omitted) or coreferentiality of matrix and subordinate subjects (too few non-coreferential sentences). Several of these factors are significant in the analysis of less common collocations. Omitting them in one analysis but not the other may skew the results. These considerations notwithstanding, the comparison of more and less common matrix subject + verb combinations discloses yet another difference between AME and QCE.

6. Connections with *like* and verbs of quotation

When circumscribing the variable context we came across problematic examples such as (8), where the word *like* appears in the same surface position as a (deletable) COMP.

8.   a.   And it seems to me [**like** if you're going to put up an ad targeting the English community you should make the effort to make sure it's grammatical.] (Donald)

     b.   Like I feel [**like** we're sort of cut off from the English community] (Donald)

Sometimes, such examples contain verb of quotation (VOQ) usages of *like* (Romaine & Lange 1991, Tagliamonte & D'Arcy 2007), as in (9a). Our labeling of these as VOQ usage of *like* stems from comparison to sentences such as (9b).

9.   a.   Because I felt **like**, "Uh: OK, I understand that the French people have to protect themselves." (Louisa)

     b.   He was **like,** "No no no no." And they were **like,** "Yes yes yes yes yes." (Liz)

Additionally, in the English data, we find an effect on the rate of COMP presence of material which intervenes between the subject and the matrix verb: there is more COMP deletion when there are intervening words (see Table 13). Specifically, we observe cases where the absence of *that* co-occurs with the presence of *like*, as in (10).

10. a.   And it seems to me <u>like</u> if you're going to put up an ad targeting the English community you should make the effort to make sure it's grammatical. (Donald)

     b.   Like I feel <u>like</u> we're sort of cut off from the English community. (Donald)

     c.   Because I felt <u>like,</u> "Uh: OK, I understand that the French people have to protect themselves." (Louisa)

Such examples raise several questions. What type of word is *like* in the canonical *feel/seem + like* + CP structure? Do these sentences contain a deleted COMP? Is *like* the COMP? Or is *like* a

conjunction? What is the relationship between *like* in the (surface) COMP position and *like* in the VOQ examples (especially *be like*, as in (9b)), given that in both cases *like* has clausal scope? Answering these questions contributes to research on the grammaticization of *like* as a COMP (*cf.* Meehan 1991, Buchstaller 2001, Meyerhoff 2002, Tagliamonte & D'Arcy 2007). We propose that the discourse marker *like* is being reanalysed as a COMP because of its frequent appearance in the same surface position as the (deletable) COMP *that*. Because *like* in its various usages has always been associated with the concept of inexact comparison, it is an ideal candidate to introduce an inexactly quoted utterance or internal thought. Our final question is: Do AMF or L1 French speakers do anything like this with the comparative French particle *comme*?

6.1.   What is like *like?*

From the perspective of grammar, it is possible that *like* is a COMP, a conjunction, or a discourse marker. First, we present grammatical and distributional evidence that *like* in this context is, in fact, a COMP. The behavior of conjunction *like* (CONJ-*like*) is distinct from COMP-*like* in several ways. The first distinction between COMP-*like* and CONJ-*like* is that CONJ-*like* can usually be replaced by *as*, another comparative conjunction (see 11a-b'), but COMP-*like* never can (11c-c').

11.   a.   Winston tastes good, **like** a cigarette should (taste good).
       a'.   Winston tastes good, **as** a cigarette should (taste good).
       b.   I feel **like** I should (feel/go to the park).
       b'.   I feel **as** I should (feel/*go to the park).
       c.   She actually feels **like** she's not an Anglophone.
       c'.   *She actually feels **as** she's not an Anglophone.
       d.   I think **like** a scientist does (think/work hard).
       d'.   I think **as** a scientist does (think/*work hard).

Second, when there is VP-ellipsis, the elided material of CONJ-*like* must find its antecedent in the conjoined clause. In (11a), *like* is unambiguously a conjunction: the elided material must come from the previous clause, but in (11b) and (11d) *like* can be a conjunction or a COMP, giving rise to two possible interpretations. *Feel like*, as in (11b), also patterns both like a conjunction and like a COMP, in allowing either interpretation.

A third reason to suppose that *like* in the relevant cases may be a COMP is that it not only licenses argument extraction, but it does so with the same asymmetry as unambiguous COMPs *that* and *for*. Objects can be extracted (as in 12a"-c") while subjects cannot (as in 12a'-c').

12.   a.   She feels **like** her friend deserves the job more.
       b.   She thinks **that** her friend deserves the job more.
       c.   She wants **for** her friend to get the job.

       a'.   *Who does she feel **like** deserves the job more?
       b'.   *Who does she think **that** deserves the job more?
       c'.   *Who does she want **for** to get the job more?

       a".   What does she feel **like** her friend deserves?
       b"   What does she think **that** her friend deserves?
       c"   What does she want **for** her friend to get?

These three grammatical arguments indicate that *like* may be interpreted as a COMP. However, we must also consider the possibility that such sentences actually have a deleted COMP and a discourse marker *like* coincidentally occurring in the same position. The frequency of null-COMP + *like*, however, is actually quite rare. There are only four examples of unambiguous cases in our corpus, listed in (13):

13.   a.   I know **like** the advertisements all focus on this worldwide thing. (Jocelyn)
       b.   I figured **like** every time: everytime the PQ says "Separation" a hundred more businesses leave. (Mike)
       c.   I think **like** toward the end like 55th, 56th, it would be about a hundred percent. (Peter)

d.   And so I remember **like** finally I said: I said "Hi" to them in French. (Donald)

Out of 256 tokens with these verbs, this is 2%. With *feel/seem*, however, 74% of the total (N = 58) are of the form *feel/seem + like*. If the *like* in *feel like* examples were a discourse marker, we would expect it to occur approximately as frequently in the same position with other verbs, but it does not. We conclude, then, that *like* is a COMP in these cases.

6.2.   Relationship with VOQ-like

There is no one-to-one correlation between using *like* as a VOQ and using it as a COMP. There are speakers who use one, the other, both and neither. These inter-speaker differences notwithstanding, it is worth considering where *like*'s use as a COMP fits into the development of the word more generally, since it has been progressively increasing its range of uses since at least Old English *soðlice* (D'Arcy 2005). Romaine & Lange (1991: 261) propose that *like* became a conjunction before branching off into a discourse marker and a VOQ. They suggest that *like* expanded from a preposition to take sentential complements, at which point it simultaneously became a conjunction and a COMP. (They do not distinguish between *like* as a conjunction and *like* as a COMP.) Vincent & Sankoff (1992) describe its use as a discourse marker.

Quotative *like*, however, is quite new on the linguistic scene: only about 30 years old (Tagliamonte & D'Arcy 2007). Discourse marker *like* in the context of a full clause, however, dates back to the early 20th century (D'Arcy 2005). Taking a clausal complement is, then, a necessary, but not sufficient precursor, and may be part of a conspiracy of multiple causes: *like*'s discourse status, which D'Arcy (2005) notes became progressively freer over the course of the 20th century, along with its subcategorization status (taking a CP complement), put *like* in a position where it could occur before a quotation of any kind.

Recall that we have examples where *feel like* introduces a quotative. We also have examples of VOQ's *you know, say,* and null verbs co-occurring with *like*:

14.   a. You know **like**: "I don't have a problem with you so you don't have a problem with
           me." (Glenn)
        b. And he said **like** "Eric this is Elizabeth." (Liz)
        c. Me **like** "Waa:" (Gloria)

The final requirement would be a pragmatic need for a new form. Tagliamonte & D'Arcy (2007: 212) propose that a rising discourse option which just preceded *like*'s appearance as a VOQ in Toronto, that of reporting inner monologue, was actually the function that VOQ *like* served at its outset of productive use. Thus, it was the combination of *like*'s COMP status, its discourse function and scope-taking ability, and a grammatical niche opening up in the context of quotation/narration strategies which all conspired to allow *like*'s introduction into the English quotative system.

6.3.   VOQ-comme

Since French *comme* functions in many similar ways as English *like*, we investigated the uses of *comme* in MAF. There are no published reports of L1 French speakers using *comme* as a VOQ.[7] MAF speakers, however, do use *comme* this way:

15.   ... *tout le monde c'est* comme *"oh tu parles à toi-même"* (Thomas)
        ... everyone is like, "Oh, you talk to yourself"

Figure 2 shows the distribution of speakers' use of *comme* in L1 and L2 Montreal French.

@@Insert Figure 2 here.

While both L1 and L2 speakers use *comme* frequently as a punctor, and share several other usages, its VOQ use is the second most common usage among L2 speakers, but is not attested in these L1 speakers. This probably reflects the L2 speakers calqueing from English, since they would not have heard it from L1 French speakers.

A natural question is why L1 French speakers would not independently have anything resembling VOQ-*comme*. Returning to the functions of *comme*, a possibility suggests itself: the conspiracy of *comme*-uses was not sufficiently strong. Earlier, we suggested that it was *like*'s multiple functions in discourse as well as its status as a COMP, along with Tagliamonte & D'Arcy's suggestion that there was an increase in use of internal monologue in discourse, that set the stage for VOQ-*like*. *Comme* is a discourse marker with considerable freedom: a conjunction, an exemplifier, an adverb and a preposition. But it does not take a clausal complement. Thus, either one or both of two facts are crucial: (1) *comme* is not a COMP, like *like* is, or (2) French speakers, though they could develop something like VOQ-*comme*, do not have the discourse need. This latter fact has not been empirically verified. If the discourse strategy of French oral narrative does not involve using internal monologue to the extent that it does in English, the necessary discourse requirements would not be present. But L2 speakers would use it, if they retain their L1 discourse strategies.

6.4. Summary for like and comme

We have suggested that *like* is a COMP with some verbs in some constructions. We provided grammatical and distributional evidence for this claim, and suggest that its COMP status might have been crucial to the development of VOQ-*like*. Finally, we examined the use of VOQ-*comme* in MAF and noted that L1 speakers did not use it contemporaneously. We suggest two possible reasons for its absence (at the time of data collection): either *comme* must take a full clausal complement (be a COMP) before it can be a VOQ, or else the grammar of French offers the possibility, but it is a possibility unrealized due to differing discourse strategies of native French speakers. Dion & Poplack (2007) report that there are now some French speakers who use this construction. This offers the exciting prospect of testing these hypotheses in the future by examining the narrative strategies of the French speakers who use it.

7. Effects of social factors

We close by assessing the roles of the social factors with this variable, trying to better understand the linguistic patterns in their social context. As this paper is part of a bigger initiative to understand the way French and English are used by Montreal Anglophones, we focus on social factors that relate to the use and meaning of COMP variation in the speech of these young bilingual Anglophones.

As one might expect, very little of the variation in the English data may be attributed to the social factors. That is, the amount and type of exposure that native English speakers have to French does not impact the way they mark their subordinate clauses in English. Of the social variables considered, the only one with a significant effect is the language used with friends and family: speakers with a Francophone family/roommate have a higher rate of COMP presence than those who do not (factor weights of 0.72 and 0.44 respectively). While this might intially be interpreted as an effect of the high rate of COMP presence in French, there is no difference between speakers with French friends and those without (both 0.44), the two other options in our scale of French contact, suggesting that the correlation seen for family/roommate may well be.coincidental. For completeness, we mention that there is no effect of sex: males and females both have factor weight of 0.50 for COMP presence in English.[8] Because this pattern is not stigmatized in spoken English, nor does it seem to be part of a change in progress, this is not surprising.

In the AMF data, one important finding from our analysis of the effects of social factors is that the more contact with native French speakers a person has, the less frequently they produce the COMP.[9] Speakers with very low contact tend to have a higher rate of COMP *que* presence, following more closely the prescriptive norm taught in school (factor weight of 0.81, vs. 0.31 and 0.47, respectively, for those with medium and high amounts of contact). This behavior resembles what we have found for pronominal variation between *nous* and *on* in French: people with very low contact tend to favor prescriptive *nous* over colloquial *on* (Blondeau et al. 2002).

This complements our analysis of several other linguistic variables (e.g. Nagy et al. 2003, Blondeau et al. 2002) in which Anglo-Montrealers who had more contact with French more closely approximated the quantitiative patterns of Montreal French than did those speakers with less contact.

8. Summary

We have shown that the systems which govern the presence of COMP in AME and AMF are quite separate, do not interact with each other, and are not obviously used as a way of marking affiliation with a particular attitude or identity. Speakers' rates of COMP presence respond to different combinations of linguistic and social factors in the two languages. And while the amount of contact that speakers have with French influences COMP rates in both languages, the effect is in the opposite direction in the two languages. The one place where we see interaction between the AME and AMF patterns is in the use of *like* and *comme* as VOQ's. This is a frequent pattern in AME and has extended into AMF in a way that native-speaker French did not exhibit at the time our data was collected. Interestingly, ten years later, its use has risen in native-speaker French to a level which is noted in at least one scholarly presentation, in a region where contact between English and French is relatively high. (Dion & Sankoff 2007). This may set the stage for this variable to gain a role in identity-marking, but that question awaits further research.

Table 1: Distribution of tokens for AME subordinate clauses

Factor group	Factor	N			%
		that	Ø	**Total**	*that*
Verb identity and frequency factors					
Lexical frequency	>1000 (in BNC)	125	539	664	18.8
	>100 but <1000	87	134	221	39.4
	<100	40	42	82	48.8
Verb identity	*realize*	15	6	21	71.4
	tell	15	8	23	65.2
	other *find*	5	3	8	62.5
	know	32	69	101	31.7
	I find	4	9	13	30.8
	say	16	39	55	29.1
	other *think*	13	75	88	14.8
	I think	38	266	304	12.5
	I'm sure	2	14	16	12.5
	I guess	4	34	38	10.5
	I wish	1	9	10	10
	I remember	3	30	33	9.1
	I don't think	5	69	74	6.8
	Other	99	84	183	54.1
Syntactic factors					
Subject type-matrix clause	1st sg.	148	618	766	19.3
	2nd sg.	8	18	26	30.8
	3rd sg. pronoun, human	15	19	34	44.1
	Pleonastic/neuter	34	15	49	69.4
	Singular noun	5	9	14	35.7
	1st pl.	3	3	6	50
	3rd pl. pronoun, human	22	20	42	52.4
	Plural noun	17	13	30	56.7
Subject type-subordinate clause	1st sg.	153	620	773	19.8
	2nd sg.	9	19	28	32.1
	3rd sg. pronoun, human	15	19	34	44.1
	3rd sg. pro. neuter	24	11	35	68.6
	Singular noun	5	9	14	35.7
	2nd plural	3	4	7	42.9
	3rd pl. pronoun, human	24	20	44	54.5
	Plural noun	19	13	32	59.4
Intervening material	none	202	645	847	23.8
	argument (IO, PP)	12	5	17	70.6
	other	38	65	103	36.9
Other *that* in COMP-position	no	245	671	916	26.7

		que	Ø	Total	%
	yes	7	44	51	13.7
Transitivity of subord. clause	intransitive	31	84	115	27
	transitive	221	631	852	25.9
Semantic factors					
Animacy-matrix subject	human	219	701	920	23.8
	pleonastic	14	5	19	73.7
	verbal	8	3	11	72.7
	abstract	11	6	17	64.7
Animacy-subord. subject	human	167	418	585	28.5
	abstract	79	291	370	21.4
	inanimate, concrete	6	6	12	50
Polarity	matrix neg., subord pos.	48	116	164	29.3
	both positive	202	597	799	25.3
	both negative	2	2	4	50
Co-referentiality	coreferential	46	140	186	24.7
	not	206	575	781	26.4
Semantic class of matrix verb	comment	21	10	31	67.7
	attitude	100	524	624	16
	extraposition	13	9	22	59.1
	knowledge	70	118	188	37.2
	utterance	38	50	88	43.2
	suasive	10	4	14	71.4
Finiteness, matrix-clause	finite	239	709	948	25.2
	non-finite	13	6	19	68.4
Phonological factor					
Segment following COMP	vowel	127	404	531	23.9
	liquid	3	6	9	33.3
	nasal	7	11	18	38.9
	fricative	105	277	382	27.5
	stop or affricate	10	17	27	37.0
Total		**252**	**715**	**967**	**26.1**

Table 2: Distribution of tokens for AMF subordinate clauses

		N			%
Factor group	**Factor**	*que*	*Ø*	**Total**	*que*
Verb identity and frequency factors					
Lexical frequency	>1000 (in Corpaix)	261	82	343	76
	>100 but <1000	97	26	123	79
	<100	29	11	40	73
Verb identity	*sembler*	5	5	10	50
	souvenir	2	2	4	50
	dire	45	19	64	70.3
	penser	87	35	122	71.3
	croire	9	3	12	75
	trouver	55	18	73	75.3
	savoir	30	9	39	76.9
	falloir	45	8	53	84.9
	voir	13	3	16	81.2
	être + X	25	5	30	83.3
	other verbs (1-4)	29	5	34	85
	other verbs (5-15)	41	8	49	83.6
Syntactic factors					
Subject type-matrix clause	1st sg.	211	77	188	73.3
	2nd sg.	5	2	7	71.4
	3rd sg.pronoun*	73	17	90	81.1
	3rd sg + 1st pl *on*	13	7	20	65.0
	2nd plural	1	0	1	100
	3rd plural	22	3	25	78.1
	Indefinite	1	0	1	100
	No subject	11	3	14	78.6
	Impersonal	25	7	32	78.1
	Noun Phrase	25	3	28	89.3
Subject type-subordinate clause	1st sg.	120	31	151	79.5
	2nd sg.	15	7	22	68.2
	3rd sg.m. pronoun	67	9	76	88.2
	3rd sg. f. pronoun	3	1	4	75
	3rd on	18	3	21	85.7
	3rd plural	23	7	30	76.7
	Indefinite	2	1	3	66.7
	Demonstrative	1	0	1	100
	No subject	8	0	8	100
	Impersonal	82	45	127	64.6
	Noun Phrase	48	15	63	76.2
Intervening material	none	329	97	426	77.2

	yes	58	21	79	73.1

Semantic factors

Polarity	positive	360	111	471	76.4
	negative	27	7	34	79.4

Phonological factors

Segment preceeding COMP	vowel	230	55	285	80.7
	[l]	7	3	10	70
	a	2	0	2	100
	[r, R]	15	5	15	75
	[z]	4	2	6	66.7
	[j]	1	0	1	100
	other fricative	41	16	57	71.9
	sibilant	82	33	115	71.3
	other stops	1	0	1	100
	[t]	3	1	4	75
	[b]	1	3	4	25
	[p]	0	1	1	0
Segment following COMP	vowel	121	19	140	86.4
	liquid	30	7	37	81.1
	nasal consonant	11	5	26	68.8
	fricative	1	1	2	50
	[j]	101	28	129	78.3
	sibilant	78	45	123	63.4
	[k]	7	1	8	87.5
	[p]	3	3	6	50
	[t]	24	6	30	80
	[d]	2	2	4	50
	[b]	9	1	10	90
Total		**387**	**119**	**506**	**76.5**

* All tokens of this type were for male pronouns.

Table 3: Profiles of the speakers

Pseudonym	Sex	Contact	Education	Friends	Work language	Grammar score
Jocelyn	F	high	high	French spouse	both	95-100%
Ted	M	high	high	French friends	both	95-100%
Vincent	M	high	high	French friends	English	95-100%
Elizabeth	F	high	high	no French friends	both	95-100%
Sandra	F	high	high	no French friends	both	95-100%
Joanie	F	high	medium	no French friends	both	85-90%
Louisa	F	medium	high	French spouse	English	85-90%
Kurt	M	medium	high	French friends	both	65-70%
Gloria	F	medium	high	no French friends	both	95-100%
Janie	F	medium	high	no French friends	English	85-90%
Glenda	F	medium	medium	no French friends	both	85-90%
Joan	F	medium	low	French spouse	both	85-90%
Greg	M	low	medium	French friends	both	75-80%
Mike	M	low	medium	French friends	both	75-80%
Jack	M	low	medium	no French friends	English	65-70%
Tony	M	low	low	French spouse	both	75-80%
Peter	M	low	low	no French friends	both	65-70%
Don	M	low	low	no French friends	English	75-80%

Table 4: Frequency counts for matrix verbs in AMF (N = 506 tokens)

Common collocation	N	% of tokens	% of tokens
Je pense 'I think'	104	21%	
Je trouve 'I find'	58	11%	
Il faut 'to be necessary'	32	6%	
C'est X 'it is X'	22	4%	
Je dirais 'I'd say'	14	3%	56%
Je sais 'I know'	12	2%	(N=283)
Je crois 'I believe'	10	2%	
Il me semble 'it seems to me'	10	2%	
Étant donné 'considering'	8	2%	
On dirait 'one would say'	6	1%	
C'est pas X 'it is not X'	7	1%	
Other *dire* 'to say'	44	9%	
Other *savoir* 'to know'	27	5%	44%
Other *falloir* 'to have to'	21	4%	(N=223)
Other *penser* 'to think'	18	4%	
other verbs, other forms of the common verbs	113	22%	

Table 5: Frequency counts for matrix verbs in AME (N = 1,013 tokens)

Common collocation	N	% of tokens	% of tokens
I think	304	30%	
I know	82	8%	
I don't think	74	7%	
I'd say, I said	34	3%	60%
I guess	38	4%	(N = 605)
I remember	34	3%	
I'm sure	16	2%	
I find	13	1%	
I figure	10	1%	
other *think*	88	9%	
other *know*	19	2%	
other *say*	22	2%	40%
tell	23	2%	(N=408)
realize	21	2%	
other verbs, other forms of the common verbs	234	23%	

Table 6: Frequency counts for matrix verbs in AMF and AME N = 967.*

Other Verbs	AME		AMF	
	N	**% of tokens**	**N**	**% of tokens**
think/ penser, croire	88 (+304)	9% (+30%)	134	26%
say/ dire	22 (+34)	2% (+3%)	64	13%
know/savoir	19 (+82)	2% (+8%)	39	8%
falloir			53	10%
tell	23	2%		
realize	21	2%		
voir			16	3%
sembler			10	2%
all other forms & other verbs	234	23%	156	31%
TOTAL	1,061		506	

*N=907 for the analysis involving lexical frequency because 60 tokens were excluded for which we were not able to establish lexical frequency of the matrix verb.

Table 7: Comparison of four analyses with 1 of 4 interacting factor groups (AME). Application value is COMP presence.

Verb identity	Factor weight	*% that*	N		
realize	0.886	*71*	21		
other *find*	0.879	*63*	8		
know	0.854	*53*	34		
tell	0.846	*65*	23		
other	0.780	*54*	183		
I said	0.680	*33*	6		
say	0.600	*35*	37		
I know	0.579	*21*	67		
I find	0.533	*31*	13		
I'm sure	0.409	*13*	16		
I think	0.365	*13*	302		
I wish	0.336	*10*	10		
think	0.320	*14*	90		
I guess	0.285	*11*	38		
I don't think	0.236	*07*	74		
I'd say	0.211	*08*	12		
I remember	0.197	*09*	33		
Semantic class				**Input**	.218
suasive	0.880	*71*	14	**Number of factors**	6
comment	0.819	*70*	30	**Range**	.289
knowledge	0.664	*37*	189	**Log likelihood**	-464
utterance	0.604	*43*	88		
attitude	0.411	*16*	624		
extraposition	0.347	*59*	22		
Finiteness				**Input**	.227
non-finite	0.783	*68*	19	**Number of factors**	2
finite	0.494	*25*	948	**Range**	.289
				Log likelihood	-464
Lexical frequency (in BNC)				**Input**	.201
>100	0.703	*43*	84	**Number of factors**	3
<100	0.614	*32*	64	**Range**	.267
>1000	0.436	*18*	539	**Log likelihood**	-415

Table 8: Binomial regression analysis for signficant factors in AME, including lexical frequency; ranked in order of decreasing significance (Application value is COMP presence, Input = .200, N=869)

Group	Factor	Weight	%
Lexical frequency (in BNC)	Frequent (>100)	0.70	43%
	Rare (<100)	0.60	33%
	Very frequent (>1000)	0.44	19%
	range	*.26*	
Matrix clause subject	Plural noun	0.82	48%
	3rd person pronoun	0.76	46%
	2nd person pronoun	0.57	31%
	Singular noun	0.49	30%
	1st person pronoun	0.46	19%
	range	*.36*	
Subordinate clause subject	1st pl.	0.66	31%
	Noun (sg. & pl.)	0.65	41%
	1st & 2nd p. sg.	0.57	26%
	Neuter & plural pronoun (human)	0.51	27%
	Non-human neuter pronoun (e.g. *it/there*)	0.50	19%
	3rd sg. pronoun (human & *'that'*)	0.22	11%
	range	*.44*	
Intervening material	argument (indirect object, prep. phrase)	0.86	71%
	other	0.62	36%
	none	0.48	22%
	range	*.38*	
Following phonological segment*	obstruent	0.61	26%
	vowel or sonorant	0.42	23%
	range	*.19*	
Polarity	both negative	0.80	50%
	matrix positive, subordinate negative	0.70	44%
	both positive	0.48	23%
	matrix negative, subordinate positive	0.48	18%
	range	*.32*	

*While this factor is significant and suggests a sonority effect, it is worth pointing out that the effect is non-monotonic when finer divisions are made in the Sonority Hierarchy. See Table 10.

Table 9: Significant linguistic factors in AMF (Application value is COMP presence, Input = 0.735, N = 506)

	Factor	Weight	%
Lexical identity of the matrix verb		.78	
	other verbs 1-4		85%
	être	.59	83%
	falloir	.58	85%
	voir	.55	81%
	other verbs 5-15	.53	83%
	trouver	.52	75%
	savoir	.50	77%
	croire	.49	75%
	penser	.43	71%
	dire	.40	70%
	souvenir	.25	50%
	sembler	.19	50%
	range	*.58*	
Following phonological environment		.65	
	vowel		86%
	liquid	.55	81%
	fricative	.51	78%
	stop	.46	77%
	sibilant	.34	71%
	range	*.31*	
Subject of the matrix clause	Noun Phrase	.71	89%
	3rd pronoun	.57	79%
	1st & 2nd pronoun	.44	62%
	range	*.27*	

Table 10: Comparison of factor weights in AME and AMF (Application value = COMP presence; For AMF: N = 506, input = 0.735; for AME: N = 967, input = 0.198)

Factor group	Factor	AMF *que*	AME *that*	
		Factor weight		
Lexical identity of the matrix verb	other common verbs (next 4 most common)	0.78	N/A	
	être	0.59	N/A	
	falloir	0.57	N/A	
	voir	0.55	N/A	
	other common verbs (next 10 most common, after other 4)	0.53	0.78*	
	savoir	0.50	0.85	*know*
			0.58	*I know*
	croire	0.49	N/A	
	penser	0.43	0.37	*I think*
			0.24	*I don't think*
			0.32	*think*
	dire	0.40	0.60	*say*
			0.68	*I said*
			0.21	*I'd say*
	souvenir	0.25	0.20	
	sembler	0.19	N/A	
Following phonological environment	vowels	0.65	0.39	
	fricative	0.51	0.66	
	liquid	0.50	0.35	
	stop	0.46	0.58	
	sibilant	0.34	0.48	
Subject of the matrix clause	NP	0.71	0.39	singular
			0.80	plural
	3rd person pronoun	0.57	0.58	feminine
			0.48	masculine
	1st & 2nd person	0.40	0.47	1st sg.
			0.23	2nd
			0.81	1st pl.

*This value is for all the uncommon English verbs combined (all tokens except the sixteen most frequent collocations and verbs).

Table 11: Comparison of factor hierarchies in L1 and L2 English and French

	QC English	AME	AMF	L1 French
Overall rate of COMP	21%	26%	77%	63%
Lexical Matrix verb	Verb identity	Verb identity/ frequency	Verb identity	Verb identity
Syntactic complexity Material between matrix verb & subord.clause	Yes > No	Yes > No	*not sig.*	Not tested
Morphological Matrix clause subject	NP >> Pronoun	NP-plural >> 3rd pronoun >> 1st & 2nd >> NP-sing (interacts w/ subord. subject)	1st & 2nd >> 3rd pronoun >> Noun (interacts w/ verb identity)	NP >> 3rd pronoun >> 1st & 2nd
Subord. clause subject	NP >> Other pronoun >> I >> it/there	*not sig.*	*not sig.*	*not sig.*
Phonological Segment after COMP	Not tested	Obs >> Son (?)	Son >> Obs	Son >> Obs
Segment before COMP	Not tested	Not tested	*not sig.*	*not sig.*

Table 12: Input values for subsamples of common vs. rare collocations in AME and QCE. Application value is COMP presence.

	Common collocations	Less comon collocations	Difference
AME	0.122	.415	0.29
QCE	0.049	.328	0.28

Table 13: Significant factor for common collocations (Analysis including lexical frequency, N = 562, Application value is COMP presence, Input = 0.127)

Intervening material	Weight	%	N
yes	0.66	23	65
no	0.47	12	497

Table 14: Significant factors for less common verbs (Analysis including lexical frequency, N = 377, Application value is COMP presence, Input = 0.435)

	Weight	%	N
Lexical frequency			
<100	0.75	68	53
>100	0.58	54	134
>1000	0.36	31	190
Subordinate subject			
Nouns	0.72	7	63
I	0.47	4	97
other pronouns	0.44	42	119
it/that/there	0.43	34	98
Intervening material			
argument	0.73	64	14
yes	0.66	64	36
none	0.47	41	327

Figure 1: Rate of COMP presence (%) in AME and AMF for each speaker
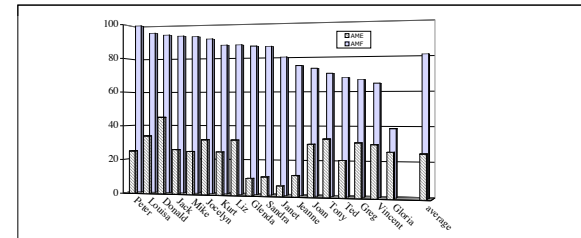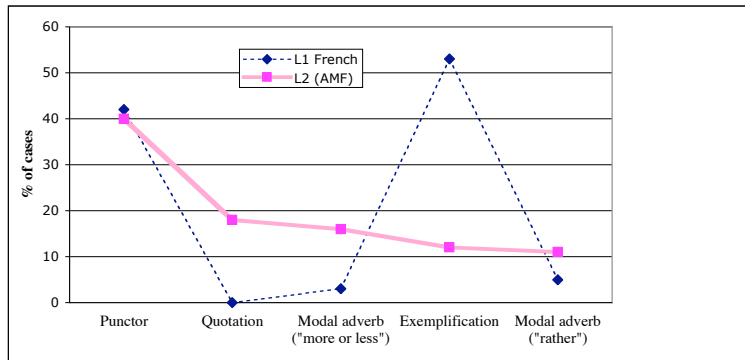
Figure 2: L1 and L2 use of comme *in Montreal French (adapted from Figure 2 in Sankoff et al.*
1997: 208)

9. Notes

[1] We list the authors alphabetically. We extend a huge thank you to Jim Wood for his contributions to coding data and organizing our arguments, particularly in the section on *like* and *comme*, and for working with us on the conference paper that preceded this version (Nagy, Blondeau and Wood 2007). We are also grateful to Miriam Meyerhoff for helpful discussions about contact-induced variation in general and subordinate markers in particular, to the interviewers Marie-Odile Fonollosa, Lucie Gagnon, and Gillian Sankoff, to the transcribers Troy Heisler, Patricia Bothner, Jim Wood, Jex Hall, Molly Mahoney, and Stephanie Buck, and especially to the speakers who generously shared their time and insights with us.

[2] This variable context was also investigated for Acadian French by Beaulieu and Chichocki (2002) and by King and Nadasdi (2006).

[3] These paragraphs on the speakers and corpus development are adapted, more or less wholesale, from Nagy et al. (2003).

[4] It is an open question how best to calculate the effect of lexical frequency (cf. Dinkin 2007, Abramowicz 2007:31-2). Is the frequency of the forms used in the particular variable context what is relevant, or does frequency of use in other syntactic positions also play a role? As a third method, we calculated the frequency of appearance for each matrix verb in the AME corpus overall (not just its uses as a matrix verb). The three types of frequency calculations are highly correlated (AME-overall & BNC: $r^2 = 0.40$, p<.05; AME-as matrix & AME overall: $r^2 = 0.28$, p<.05). Further inquiry might reveal which method of frequency calculation best fits the model, but we proceed using the BNC calculations, which contribute a strong effect for lexical frequency in the multivariate analyses presented below.

[5] We could also compare our results with Warren (1984), which identified a rate of 88% COMP presence. Since her circumscription of the variable context is slightly different, we prefer to compare our results with Dion (2003) who defined the variable context as we did.

[6] The expression *disons* 'let's say' is not frequent enough in our data set (less than 5 occurrences) and is not further discussed. The expression *parce que* is not part of the variable context and is also not further discussed.

[7] Dion and Poplack (2007) recently reported the use of VOQ-*comme* to be surprisingly common in the French spoken by students in the Gatineau area of Quebec.

[8] This was determined using the step-up/step-down analysis in Goldvarb, including all the social factors and the linguistic factors which proved significant in the analyses presented above.

[9] This was the only social factor selected as significant in the multivariate analysis.

References

Abramowicz, Lukasz. 2007. Sociolinguistics meets Exemplar Theory: Frequency and recency effects in (ing). *U. Penn Working Papers in Linguistics* 13(2): 15-6.

Beaulieu, Louise and Chichocki, Wladislaw. 2002. Le concept de réseau social dans une communauté acadienne rurale. *Revue canadienne de linguistique* 47(3): 123-150.

Berkenfield, Catie. 2001. The role of frequency in the realizeation of English *that*. In *Frequency and the Emergence of Linguistic Structure,* Joan Bybee (ed.), 281-308. Philadelphia: John Benjamins.

Biber, Douglas. 1999. A register perspective on grammar and discourse: variability in the form and use of English complement clauses. *Discourse Studies* 1: 131-150.

Blanche-Benveniste, Claire. 2000. Corpus de français parlé. In *Corpus: Méthodologie et applications linguistiques,* Mireille Bilger (ed), 123-8. Paris: Champion et les Presses Universitaires de Perpignan.

Blondeau, Hélène, Nagy, Naomi, Sankoff, Gillian and Thibault, Pierrette. 2002. La couleur locale du français des Anglo-Montréalais. *In L'acquisition de la variation par les*

apprenants du français langue seconde. [*Acquisition et Interaction en Langue Étrangère* 17] Jean-Marc Dewaele and Raymond Mougeon (eds.), 73-100.

Buchstaller, Isabelle. 2001. An Alternative view of *like*: Its grammaticalisation in American English and beyond. *Edinburgh Working Papers in Applied Linguistics* 11(1): 21-41.

Connors, Kathleen. 1975. L'effacement de *que* - règle syntaxique. *Recherches linguistiques à Montréal* 4: 17-33.

D'Arcy, Alexandra. 2005. Like: Syntax and Development. Ph.D. dissertation, University of Toronto.

Dinkin, Aaron. 2007. The real effect of word frequency on phonetic variation. Paper presented at Penn Linguistics Colloquium 31.

Dion, Nathalie. 2003. L'effacement du *que* en français canadien: Une étude en temps réel. Master's thesis, University of Ottawa.

Dion, Nathalie and Poplack, Shana. 2007. Linguistic mythbusting: The role of the media in diffusing change. Paper presented at NWAV 36, University of Pennsylvania.

Gadet, Françoise. 2007. *La variation sociale en français.* 2nd edition. Paris: Orphys.

King, Ruth and Nadasdi, Terry. 2006. Another look at *que* deletion. Paper presented at NWAV 35, Ohio State University.

Leech, Geoffrey, Rayson, Paul and Wilson, Andrew. 2001. Companion website for Word Frequencies in Written and Spoken English: based on the British National Corpus. Longman: London. http://www.comp.lancs.ac.uk/ucrel/bncfreq/lists/2_2_spokenvwritten.txt.

Martineau, France. 1985. Élision variable de (*que*) dans le parler d'Ottawa-Hull. Master's thesis. Université d'Ottawa.

Meehan, Teresa. 1991. It's like, "What's happening in the evolution of 'like'?": A theory of grammaticalization. *Kansas Working Papers in Linguistics* 1: 37-51.

Meyerhoff, Miriam. 2002. All the same? The emergence of complementizers in Bislama. In *Reported Discourse: A meeting ground for different linguistic domains*, Tom Güldemann and Manfred von Roncador (eds.), 341-359. Amsterdam: John Benjamins.

Nagy, Naomi and Blondeau, Hélène. 2005. I think [CP [COMP Ø] [IP *that* is deleted in Montreal]]. Poster presented at NWAV 34, New York University.

Nagy, Naomi, Blondeau, Hélène and Auger, Julie. 2003. Second language acquisition and "real" French: An investigation of subject doubling in the French of Montreal Anglophones. *Language Variation and Change* 15.1: 73-103.

Nagy, Naomi, Blondeau, Hélène and Wood, Jim. 2007. On était comme, "We think 'like' may be a COMP too." Paper presented at NWAV 36, Philadelphia.

Poplack, Shana. 1989. The care and handling of a megacorpus: the Ottawa-Hull French project. In *Language Change and Variation*, Ralph Fasold and Debra Schiffren (eds.), 411-451. Amsterdam: John Benjamins.

Poplack, Shana. 2001. Variability, frequency, and productivity in the irrealis domain of French. In *Frequency and the emergence of linguistic structure*, Joan Bybee and Paul Hopper (eds.), 405-428. Amsterdam: John Benjamins.

Poplack, Shana, Walker, James and Malcolmson, Rebecca. 2006. An English like no other? Language contact and change in Quebec. *Canadian Journal of Linguistics* 51(2): 185-213.

Romaine, Suzanne and Lange, Deborah. 1991. The use of 'like' as a marker of reported speech and thought: A case of grammaticalization in progress. *American Speech* 6: 227-279.

Sankoff, David, Tagliamonte, Sali and Smith, Eric. 2005. Goldvarb X: A variable rule application for Macintosh and Windows. Department of Linguistics, University of Toronto.

Sankoff, Gillian. 1980. *The Social Life of Language*. Philadelphia: University of Pennsylvania Press.

Sankoff, Gillian and Brown, Penelope. 1976. The origins of syntax in discourse: a case study of Tok Pisin relatives. *Language* 52: 631-66. Reprinted in *The Social Life of Language,* Gillian Sankoff (ed.), 211-255. Philadelphia: University of Pennsylvania Press.

Sankoff, Gillian, Sarrasin, Robert and Cedergren, Henrietta. 1971. Quelques considérations sur la distribution de la variable *que* dans le français de Montréal. Paper presented at the Congrès de l'Association canadienne-française pour l'Avancement des Sciences.

Sankoff, Gillian, Thibault, Pierrette, Nagy, Naomi, Blondeau, Hélène, Fonollosa, Marie-Odile and Gagnon, Lucie. 1997. Variation and the use of discourse markers in a language contact situation. *Language Variation and Change* 9: 191-218.

Tagliamonte, Sali and D'Arcy, Alexandra. 2007. Frequency and variation in the community

grammar: Tracking a new change through the generations. *Language Variation and*

Change 19: 199-217.

Thibault, Pierrette and Daveluy, Michelle. 1989. Quelques traces du passage du temps dans le

parler des Montréalais, 1971-1984. *Language Variation and Change* 1(1): 19-45.

Thompson, Sandra and Mulac, Anthony. 1991. The discourse conditions for the use of the

complementiser *that* in conversational English. *Journal of Pragmatics* 15: 237-251.

Torres Cacoullos, Renee & Walker, James. 2003. What's *that*? Paper presented at the Canadian

Linguistic Association. Dalhousie University, Halifax, NS.

Torres Cacoullos, Renee Walker, James. *forthcoming*. On the persistence of grammar in

discourse formulas: A variationist study of *that*. *Linguistics* 46(6).

Véronis, Jean. 2000. Fréquences des mots en français parlé. http://www.up.univ-

mrs.fr/~veronis/donnees/.

Vincent, Diane and Sankoff, David. 1992. Punctors: A Pragmatic variable. *Language Variation*

and Change 4: 137-62.

Warren, Jane. 1994. Plus ça change, plus c'est pareil: The case of 'que' in Montreal French.

Culture 14(2): 39-49.

Appendix A: Examples for each factor code, in AME

This section provides illustrative examples from our corpus for each factor in all of the linguistic

factors considered. Factors are also defined, as necessary.

Matrix verb identity and frequency
Lexical frequency of the matrix verb

For the English data, we used Leech et al. (2001), which provides lemmatized frequencies for

spoken British English, from the British National Corpus.

very frequent (>1000 in BNC)	That means [we won't have our jobs.] (Victor)
frequent	(>100) I mentioned [I love uh geography] (Paul)
infrequent	(>10) I would imagine [he didn't.] (Paul)
rare	(<10) It didn't register [that that it was very important] (Paul)
other (not listed in the ANC)	She's just absolutely amazed [that an entire culture speaks two languages.] (Lisa V)

Matrix verb identity

Common collocations

I think	I think[it was only two people.] (Elizabeth)
I don't think	I don't think [it's a big problem.] (Sarah)
I guess	I guess [I wasn't persecuted enough.] (Elizabeth)
I remember	I remember [the teacher was always on my back.] (Victor)
I find	Now I find [they're very judgemental up there] (Terry)
I'm sure	I'm sure [it's helpful] (Elizabeth)
I wish	I wish [that I could do that.] (Doug)

Other forms

realize	That's when I I realized [that French was all around me] (Paul)
other forms of *find*	I found [I'd get home] (Matt)
tell	He tells you [well Principal Black might help us.] (Victor)
know	I know [that it's wrong]. (Lisa V)
say	My parents said [that was a dumb move,] (Matt)
remind, other forms of *remember*	reminds them [that I do.] (Elizabeth)
thought, other forms of *think*	I thought [I'd lost it.] (Sarah)
figure	I figure [it should be a business choice.] (Matt)

other	I felt [I was forgetting a bit] (Paul)

Syntactic factors

Subject type (matrix clause)

1st sg.	I know [they speak 2 languages at home.] (Victor)
2nd sg.	You said [you spoke to him.] (Paul)
3rd sg. f. pronoun	She's just amazed:[that we speak both languages.] (Lisa V)
3rd sg. m. pronoun	He finds [that I'm a completely different person.] (Sarah)
3rd sg. Noun Phrase	My mum tells me [I used to come home in tears.] (Sarah)
1st pl.	We never thought [it was important.] (Glenn)
3rd pl. pronoun	They could detect [that I I was an Anglophone] (Paul)
3rd pl. Noun Phrase	People are aware [that they'll freak out at you, if you speak in English.] (Karl)
neuter / pleonastic	It doesn't mean that [the class will not be conducted in English.] (Johanna)
other (e.g. gerund, infinitive)	I would hate to think [that it was that]. (Lisa V)

Subject type (subordinate clause)

1st sg.	I felt [I was forgetting a bit] (Paul)
2nd (sg./pl.)	I don't imagine [you got along with every one of your teachers.] (Paul)
3rd sg. f. pronoun	I just thought [she was such a pixy.] (Lisa V)
3rd sg. m. pronoun	I would imagine [he didn't.] (Paul)
3rd sg. Noun Phrase	I thought [French im: immersion worked really well.] (Lisa V)
1st pl.	They think [we talk behind their back and whatever.] (Victor)
3rd pl. pronoun	I never thought [they were like that.] (Victor)
3rd pl. Noun Phrase	Cause I know [some people don't.] (Elizabeth)
neuter / pleonastic	Cause I know [there's a lot of politics involved.] (Victor)

Material between COMP and subordinate clause

intervening clauses, fillers, parentheticals	I suspect maybe [that if you go through the French immersion,] (Terry)
argument	I'd explain it to them [that in Westmount it was still Dorchester.] (Doug)
pause	So I think [that: that: that really was the case.] (Jeff)
none	I think [it's just north] (Jeff)

Non-COMP *that* in COMP-position

none	I don't think [it's gonna happen again.] (Karl)
yes	But I don't think [Ø that's gonna support four million people.] (Karl)

Type of matrix clause verb

non-finite	I would like to believe [that it's possible.] (Elizabeth)
finite	I don't think [he could function well.] (Paul)

Transitivity of subordinate clause

transitive	They think, you know, [we're gonna lose a distinct society and stuff.] (Karl)
intransitive	I don't think [they really care] (Gabrielle)

Phonological factor

Phonology of segment following COMP

[+sonorant]

nasal	Cause I knew [that my French wasn't that good.] (Doug)
liquid	But I know [last year she was in Australia.] (Jenny)
vowel	I figure [I'll respect that in class] (Terry)

[-sonorant]

affricate	So I figured [just be mov- easier to move elsewhere] (Gail)
stop	Because I think [both groups called the other group Pepsi.] (Doug)
fricative	They were told [that the classes were full]. (Lisa V)
sibilant	Cause I know [some people don't.] (Elizabeth)

Semantic factors

Animacy of subject in matrix clause

human	They never think [I'm from another part of Quebec] (JohaneB)
abstract	The problem is [they don't have the heart for it.] (Victor)
pleonastic	It seemed [like there was a lot.] (Gail)
no subject (e.g. gerund, infinitive)	That person's English too, thinking [I'm French.] (Karl)

Animacy of subject in subordinate clause

human	I don't think [he could function well.] (Paul)
concrete inanimate	He'd tell me [that this bathroom floor wasn't washed]. (Matt)
abstract	I said [her vocabulary and everything was great] (Elizabeth)
pleonastic	I think [there is a chance.] (Elizabeth)
gerund, infinitive	I don't think [that uh separating's going to give them any more of an identity.] (JohaneB)

Polarity

matrix positive, subord positive.	It just seems [they all have this attitude] (Terry)
matrix negative, subord negative	My mother won't admit [she can't speak French.] (Doug)
matrix negative, subord positive.	But it doesn't mean [you know what it is.] (Doug)
matrix positive, subord negative	I felt [that the boss wasn't too pleased] (Paul)

Co-referentiality

coreferential	I'm just glad [I learned both] (Terry)
non-coreferential	I realized [that French was all around me] (Paul)

Semantic class of the matrix verb

comment/factive	I'm just glad [I learned both] (Terry)
extraposition	It just seems [they all have this attitude] (Terry)
suasive	He kept insisting [that he was Trinidadian] (Matt)
knowledge	I knew [it was school.] (Jenny)
utterance	You said [you spoke to him.] (Paul)
attitude	I would hate to think [that it was that]. (Lisa V)

Appendix B: Examples for each factor code, in AMF

This section provides illustrative examples from our corpus for each factor in all of the linguistic factors considered. Factors are also defined, as necessary.

Matrix verb identity and frequency

Matrix verb frequency

For the French data, we used Véronis (2000), which provides lemmatized frequencies for spoken French, based on Corpaix, the 2000 version. For a description of this corpus of spoken French, see Blanche-Benveniste (2000).

>1000 (in Corpaix)	Ça se peut [qu'elle parle anglais.] (Ted)
>100 but <1000	Je rappelle [quand j'étais toute petite tout en anglais.] (Janie)
<100	Je perçois facilement [que la personne est anglophone.] (Tony)

Matrix verb identity

sembler	Il me semble [que toutes ces écoles-là leur programme français c'est pas adéquat.] (Louisa)
souvenir	Je me souviens [que ça avait été bizarre une expérience vraiment étrange.] (Elizabeth)
dire	On dit [qu'il y a beaucoup de drogue.] (Jack)
penser	Je pense pas [que je vais finir le programme.] (Janie)
croire	Je crois [que ça avait affaire avec la loi.] ((Joanie)
trouver	Moi je trouve [c'est une question de culture.] (Ted)
savoir	Durant le deuxième terme je savais [c'était pas pour moi.] (Joanie)
falloir	Tu sais faut [que je pense avant de parler.] (Jack)
voir	Alors j'ai vu [que c'est inutilisable en Quebec.] (Jack)
être + X	C'est juste bon [que le monde là-bas chu : c'est les Blocks là les vrais.] (Ted)
Other verbs (1-4)	Ma mère exige [que je travaille.] (Ted)
Other verbs (5-15)	Je remarque [qu'il y en a beaucoup plus.] (Janie)

Syntactic factors

Subject type (matrix clause)

1st sg.	Je ne crois pas [qu'il y a discrimination dans l'emploi.] (Jack)
2nd sg.	Tu veux [que je dise qu'est-ce qu'il a fait?] (Kurt)
3rd sg.	Il faut [que je sois capable de réfléchir comme lui.] (Elizabeth)
3rd sing + 1st pl. *on*	On dirait [que tu es comme ça tout le temps.] (Ted)
2nd pl.	Vous saviez [que j'étais un anglais.] (Ted)

3rd pl.   Ils sauraient [qu'elle aurait de la misère.] (Ted)   Negative   Faut pas dire [que les bons professeurs ils font ça.] (Ted)
Indefinite   C'est pas [qu'ils ont pris une décision.] (Kurt)
No subject   Étant donné [tu peux pas répondre à la demande de tous.] (Vincent)
Impersonal   Personne dans mon famille peut comprendre rien [qu'il a dit.] (Don)
Noun Phrase   Ma femme a décidé [que il y a beaucoup de monde.] (Don)

Subject type (subordinate clause)
1st sg.   Je trouve [que je suis plus fort pour faire les courses.] (Don)
2nd sg.   Je dirais [quand tu fais des recherches sur des costumes.] (Sandra)
3rd sg. m.   Vu [qu'il avait commencé dans l'école anglophone.] (Sandra)
3rd sg. f.   Elle a vu [qu'elle était très responsable.] (Victor)
3rd sing + 1st pl. *on* Mes parents ont voulu [qu'on aille à l'école en français.] (Sandra)
3rd pl.   Ces gens je pense [qu'ils on trouvé.] (Peter)
Indefinite   Je pense pas [que c'était tant que ça.] (Sandra)
No subject   Je pense [que oui.] (Kurt)
Impersonal   Je pense [que la personne avec qui ils ont fait des enfants.] (Tony)
Noun Phrase   J'ai trouvé [que l'allemand c'était plus facile d'apprendre.] (Peter)

Material between COMP and subordinate clause
Intervening material   Ca se peut très facilement [que la première fois que je parle
                       anglais.] (Louisa)
No material   Je le sens [que je danse pas pareil comme eux.] (Victor)

Phonological factors

Phonology preceding COMP
      Vowel   C'est pas [qu'ils vont tuer.] (Victor)
      Liquid   Je rappelle [quand j'étais petite tout en anglais.] (Janie))
      Fricative   Moi je trouve [c'est ça c'est les jeunes.] (Ted)
      Stop   Je remarque [qu'il y en a beaucoup plus.] (Janie)
      Sibilant   Je pense [la plupart sont francophones.] (Jack)

Phonology of segment following COMP
      Vowel   J'ai remarqué [on a des plus par exemple.] (Victor)
      Liquid   Je trouve [que la famille noire est plus solidaire.] (Victor)
      Fricative   Je trouve [faut le faire.] (Joan)
      Stop   Je pense [que quand j'avais : quand on m'a donné le char il y avait un
                  couvre-feu.] (Jack)
      Sibilant   Je savais [que c'était gagné avec elle.] (Victor)

Semantic factor

Polarity
      Positive   Il faut [qu'il travaille dans deux langues.] (Mike)