I think (that) something's missing: Complementizer deletion in nonnative e-mails

Mercedes Durham
University of Aberdeen
mercedes.durham@abdn.ac.uk

Abstract

Sociolinguistic competence is not often examined in nonnative English acquisition. This is particularly true for features where the variants are neither stylistically nor socially constrained, but rather are acceptable in all circumstances. Learning to use a language fully, however, implies being able to deal with this type of 'difficulty,' and understanding what type of variable features nonnative speakers acquire with ease and which ones they do not may help us better understand more general processes of second language acquisition. By comparing the rates of complementizer deletion of nonnative to native speakers and examining their distributions across various internal and external factors, this paper addresses these issues and offers an example of acquisition of what is, in some ways, an invisible variant. Furthermore, by focusing on a Swiss student association, the paper is also able to compare the patterns of French, German and Italian native speakers, to examine to what extent they differ in English.

Keywords: sociolinguistic competence, complementation, complementizer deletion, zero complementizer

Learning how to appropriately use the syntax of a new language is never an easy task, but it can further be complicated in cases where the target language demonstrates variation within a single construction – be this variation linked to social, stylistic or linguistically internal factors (see Bayley, 1995; Bayley & Regan, 2004; Mougeon, Nadasdi, & Rehner, 2010; Mougeon & Rehner, 2001;

Mougeon, Rehner, & Nadasdi, 2000; Regan, 1995; Regan, Howard, & Lemée, 2009; Rehner, Mougeon, & Nadasdi, 2003 for examples of this). How do nonnative speakers cope with learning aspects of grammar that not only are variable, but where one of the variants is actually not there? In the case of complementizers in English, speakers, native and nonnative, have the option to use *that* or simply to use a zero form (Examples (1)-(2)) in most cases.

(1)   *I hope Ø you enjoyed the day and liked the city.* (b, Italian, e-mail)[1]
(2)   *I hope that in the future in all Switzerland we'll have some common projects at national level.* (f, French, e-mail)

Native speakers, as we shall see, 'decide' whether to use the overt (*that*) or covert (Ø) complementation form according to a number of factors, internal and external, but what of nonnative speakers? Are they able not only to use both variants, but crucially use them in the same way as native speakers and at equivalent rates? Will it depend on whether the structure is similarly optional in their source language? Or do nonnative speakers simply avoid the covert variant, given that the use of *that* is grammatically, though not pragmatically, correct in all circumstances?

Complementizer variation can offer valuable insight into nonnative language acquisition processes in a number of ways; first of all, because it has been extensively studied in native varieties of English and as such will make it easier for us to establish whether nonnative speakers have acquired the same patterns; secondly, because the variation is restricted to two variants both of which are possible in all situations; and finally, because it is a feature that is not generally taught or even pointed out in teaching, we will be able to see whether nonnative speakers acquire aspects of the target language which, in some ways, are 'not there.' In addition, this analysis will focus on the speech of Swiss natives, which will provide additional insight into strategies of acquisition, as it will enable us to compare French, German and Italian native speakers.

The section below will introduce studies which have examined the nonnative acquisition of aspects of sociolinguistic competence – that is to say what is variable in native speech – and summarize what is already known in terms of nonnative speakers' ability to acquire the nuances of native language use. It will be followed by a discussion of complementizers in English and then a presentation of the corpus analyzed and, because the speakers in the corpus are from

---

[1] The codes for the examples from the data are as follows: the one letter code given to the speaker in the corpus, their native language and the medium the example was delivered in.

Switzerland, a brief portrait of the situation of English in the country. The results and a discussion will come before a general conclusion.

Acquisition of Sociolinguistic Competence

The examination of variable features of the target language in nonnative speech is a relatively new strand of Second Language Acquisition research, and has, in many ways, chosen to distinguish itself from "the bulk of previous research in Second Language Acquisition (SLA) [which] focused on aspects of the target language where native speakers display invariant language usage (i.e., use only one linguistic element to convey a given notion)" (Mougeon & Rehner, 2001, p. 398). Indeed this "new strand of research" is quite different from traditional research into second language acquisition in that it "includes not only factors that have been examined by mainstream SLA research, but also those that have been found to be correlated with L1 variation in sociolinguistic research" (Mougeon & Rehner, 2001, pp. 398-399). This type of research looks for proof that L2 learners can show "the same kind of sociolinguistic ability in using the variants as do L1 speakers (i.e., ability to observe the linguistic and extralinguistic constraints that have an impact on variant choice)" (Mougeon & Rehner, 2001, p. 399).

Basically, if nonnative speakers are to be considered to have fully mastered the target language, they must show they have acquired the syntactic and phonological aspects of it. The nonnative speakers will also have to display that they have acquired the variable rules of native speakers, both for features where the variation is stylistically motivated and those where the variation is internally constrained. These variable rules belong to native speakers' communicative competence (Hymes, 1972, p. 281) and are an intrinsic part of the mastery of one's own language.

A number of studies concerned with this aspect of acquisition have shown that sociolinguistic competence is not always fully acquired (see for example Dewaele, 2004; Dewaele & Regan, 2002; Regan, 1995; Rehner, Mougeon, & Nadadsi, 2003). It has been found, for example, "that immersion students learn an academic register of the L2 but not its vernacular" (Lyster, 1996, p. 167). This is because "whereas there exist numerous dictionaries and reference grammars to support the teaching of lexis and syntax, there are no such reference books to support the teaching of sociolinguistic variation" (Lyster, 1996, p. 167).

It is not surprising that reference books do not tend to present sociolinguistic variation, as, for the most part, even native speakers are not aware of the variable rules they use everyday. The fact that the application of these variable rules is almost completely subconscious for native speakers means that they may be

more difficult for nonnative speakers to notice and acquire. The difficulties experienced by the students in these studies are not restricted to external constraints but to internal factors as well (as demonstrated in Regan, 1995).

Rehner, Mougeon and Nadasdi (2003) and Mougeon, Nadasdi and Rehner (2010) examined nonnative French speakers in immersion classrooms in Canada, while Nagy, Blondeau and Auger (2003) and Sankoff, Nagy, Blondeau, Fonollosa and Gagnon (1997) studied other speakers of nonnative French in Canada. Aspects of nonnative French acquisition have also been studied in Europe (Dewaele & Regan, 2002; Regan, 1995; Regan, Howard & Lemée, 2009). Bayley has conducted a considerable amount of research on the variation patterns of Chinese speakers, either learning Chinese as a heritage language in the United States, or learning other languages abroad (Bayley, 1995; Langman & Bayley, 2001). Relatively little has been done on the nonnative acquisition of English in terms of sociolinguistic competence (see Durham, 2007 for a further discussion of this), so this paper provides a first examination of this topic.

This discussion should not be taken to imply that other types of second language research have not considered related aspects as well. Indeed, a number of papers from recent years have focused on examining why there is "a disjunction between success in acquiring the syntax of the target language (TL), on the one hand, and persistent difficulties at the interfaces of syntax with other grammatical modules, e.g. discourse-pragmatics, on the other" (Hopp, 2007, p. 147; see also Sorace & Filiaci, 2006). Of course, interface structures have not always been found to pose a difficulty for nonnative speakers and a comparison of various studies examining interface issues has lead White (2010) to underline that "interfaces are not monolithic: it is not the case that all interfaces lead to difficulties, it is not the case that all phenomena at a particular interface are necessarily problematic, it is not the case that acquisition failure is inevitable" (p. 11).

Complementizers

Syntactic structures, of the sort introduced with verbs such as *think*, *say*, and *mean* can either have a *that* complementizer between the verb and the following clause or zero (as in Examples (1) and (2) above). Far from being free variation, the use of the two variants is constrained by a range of factors.

Two very general findings related to this were reported in previous studies. First of all, the use of the zero form has grown through the history of English. Although the zero form was rarely used in Old and Middle English, by the twentieth century it had achieved a near categorical use for some specific

verbs (Tagliamonte & Smith, 2005, p. 301; Thomson & Mulac, 1991, p. 244; Torres Cacoullos & Walker, 2009, p. 2). Secondly, Elsness (1984, p. 521) found that in formal writing zero complementizers are used far less than in informal writing, so it would be said that, in present day English, style exerts a considerable effect on the selection of zero complementizer forms.

Looking at informal oral data, Thomson and Mulac (1991, p. 242) found that overall the zero complementizer was used at a rate of 86%, Tagliamonte and Smith (2005, p. 300) found a rate of 84%,[2] Torres Cacoullos and Walker (2009, p. 16) found 79% and Kolbe (2008, p. 112) found 90%.[3] Elsness (1984, p. 521), examining a subsection of written data from the Brown corpus, found rates between 52% and 58% in his two informal text categories (Press Releases and Fiction: Adventure & Western) and far lower rates (between 1.3% and 15%) in his two formal text categories (Belles Lettres & Biography and Learned & Scientific Writing) (see Kucera & Francis, 1967 and Ellegard, 1978 for further information about the Brown corpus and the Syntax Data Corpus which is a subset of it).

Nonnative speakers from a variety of linguistic backgrounds use both variants, as can be seen in Examples (3)-(8) below.

French speakers
(3) *I think* that *there's a virus in the past document I send to you.* (f, French, e-mail)
(4) *I will take some copies [. . .] because I think Ø there is a virus.* (f, French, e-mail)

Italian speakers
(5) *I guess* that *there will not be very many new people.* (b, Italian, e-mail)
(6) *I guess Ø you're back in Switzerland!* (b, Italian, e-mail)

German speakers
(7) *So I also think* that *we need to think carefully about the division of expenses.* (h, German, e-mail)
(8) *I think Ø it is very important to have such a useful booklet.* (h, German, e-mail)

---

[2] Tokens such as *I think, you know* and *I mean,* which categorically selected the zero complementizer, were excluded in Tagliamonte and Smith's results.
[3] Note, however, that what was included in the analysis did vary from study to study. Kolbe chose to examine solely the verbs *think, say, know* and *see,* which goes some way in explaining why her rates were higher than in the other studies.

A survey of several grammar books used in Switzerland to teach English reveals that the existence of two variants is never explicitly made clear. The two forms are used, however, both in the grammar books (e.g., Soars & Soars, 1987; Spencer, 1999) and other teaching materials[4] the Swiss students use and in the speech of their teachers. If Swiss speakers have the same patterns as native speakers, we can hypothesize that they acquired these patterns subconsciously then and not through overt and conscious teaching.

What of complementizer forms in the native languages? French and Italian do not have a zero complementizer variant. Complementizer forms in these two languages are somewhat similar to relative pronouns (for a further discussion, see Durham, 2007, p. 146-148), in that the complementizer particle is also the more frequent of the possible relative pronouns (*que* for French, and *che* for Italian; Examples (9)-(10)). German and Swiss German, however, have both overt and zero complementizer forms (Examples (11)-(12)). Moreover, similarly to English, the *dass* form is seen to be more formal than the zero form. The variable patterns of complementizers have not been studied in German, however, so we cannot know if some of the other factors found to be significant in English complementizer use operate in German as well.

(9)   *Je pense* que *tu as presque fini.*
(10)   *Penso* che *hai quasi finito.*
(11)   a. *Ich glaube,* dass *du schon fertig bist.*
        b. *Ich glaube* Ø *du bist schon fertig.*
(12)   *d Ruth glaubt,* dass/Ø *d Susann het s gmacht* (Penner and Bader, 1995, p. 103).

Native English Studies

Studies which have focused on contemporary varieties of English (Elsness, 1984; Kolbe, 2008; Tagliamonte & Smith, 2005; Thompson & Mulac, 1991; Torres Cacoullos & Walker, 2009) can help us establish how close the nonnative speakers are to native norms.

Tagliamonte and Smith (2005) provided an in-depth presentation and summary of earlier studies and attempt to establish which of the factors mentioned in these studies are relevant to their own data. Their analysis focused on the patterns of zero complementizer use of the oldest generation of speakers in several relatively isolated northern British communities (in Cumbria,

---

[4] Much of the literature read by students of English would have contained instances of both *that* and zero complementizer.

Lowland Scotland and Northern Ireland). They found very high percentages of the zero complementizer (around 90%) and established that the zero form had become nearly categorical in contexts such as *I think* and *I mean (*Tagliamonte & Smith, 2005, p. 299). In the variable contexts of use, they also uncovered a number of internal factors which conditioned the use of the variants: first and second person subjects versus other subjects, present tense versus past tense, the presence of additional elements in the verb phrase and finally the use of adverbials between the verb phrase and complementizer (Tagliamonte & Smith, 2005, p. 301). These factors will be examined in this study and presented in more detail in the section on extraction and coding.

Moreover, as mentioned above, formality also exerts a considerable influence on the use of complementizer variants. All the studies mentioned here which examined oral data found high rates of deletion. On the other hand, Elsness (1984, p. 521), who examined written data, found low rates of the zero form in the more formal texts. He found higher rates of the zero form in the less formal texts but these rates were still rather lower than the oral data in other studies. This is particularly relevant, because this study examines e-mail data, as will be discussed more fully directly below. The place of e-mail on the continuum between oral and written data is debatable (cf. Herring, 2001), so we cannot know a priori if the rates for the zero complementizer in e-mail data will be high, as in spoken English, or low, as in formal written English. A native e-mail control corpus will be examined as well, and will play a crucial role here as it is directly comparable to the nonnative data. If the native e-mail data shows high levels of zero complementizer, then we would expect high levels from the nonnatives as well, if we are to prove that the non-native speakers pattern like native speakers. On the other hand, if the native data shows low levels of deletion, then we would expect the same from the nonnatives.

Data

In many respects, Switzerland is a perfect place to examine nonnative English acquisition; not only does the mix of French, German and Italian speakers mean that it is possible to examine effects of various source languages at once, but, furthermore, Switzerland is on the cusp of transitioning from an English as a Foreign Language country (or Expanding Circle country in Kachru's terminology; 1982, p. 38) to an English as a Second Language country (Outer Circle) by virtue of the fact that English is used as an intranational lingua franca in a number of domains (see Durham, 2007 and Dürmüller, 2001, 2002 for a further discussion of this). This means that many of the users of English in Switzerland have a high competency in the language.

This is the case of the Swiss speakers in the present study; the corpus of data is composed of a collection of 653 e-mails (circa 90000 words) sent over a period of 4 years by medical students who were all members of the International Federation of Medical Students' Associations – Switzerland (IFMSA-CH) (see Durham, 2003 for a full discussion of the mailing list and the linguistic background of its members). The association is composed of students who are studying at the various medical schools in Switzerland at the universities of Lausanne, Geneva, Berne, Zurich and Basel (IFMSA, 2003; IFMSA-CH, 2003) and whose linguistic backgrounds are French, German or Italian.[5] As described by one of the members of IFMSA-CH, the purpose of the association "*is to enable international cooperation in professional training and the achievement of humanitarian ideals*" (b, Italian, e-mail). The association's main use of English is in e-mails and, at the time of data collection, some members e-mailed on a daily basis, so it is the natural place to examine their English use.

Because the main part of the Swiss medical student data was composed of e-mails, a comparable corpus of e-mails sent by native English-speaking British students was collected as well and will be the main point of comparison for the Swiss data (again, see Durham, 2007 for further discussion of this control data set).

Extraction and Coding

Every instance where either *that* or zero could have been used was extracted from the data and was coded for a number of factors.[6] The factors which were coded in this analysis and which will be studied in detail are very similar to those examined in Tagliamonte and Smith (2005) and Torres Cacoullos and Walker (2009). As well as *speaker* and *speaker's native language*, which are the external factors in this analysis, I will focus on *subject of the main clause*, the *tense of the verb*, whether there are any *additional elements to the main clause*, whether the verb phrase and the complementizer are *separated by adverbs or adverbials* and finally the *lexical verb* which the complementizer follows.

---

[5] Although some students were slightly more proficient in English than others, they had all had a similar amount of schooling in English. Furthermore, only one speaker on the mailing list was an English-speaking bilingual and was excluded from the analysis. Note furthermore that most of the students had also studied at least one of the other Swiss languages, which possibly makes them rather different linguistically from German, French and Italian speakers of English in Germany, France and Italy respectively.

[6] Rather differently from Tagliamonte and Smith (2005), there were no cases of parentheticals which were in a position where the *that* form was not a possible option (*e.g.*, *She's very nice, I think*).

Previous research on complementizers found first and second person subjects have a higher proportion of zero complementizers than third person subjects. The tokens have been coded for whether they are first person singular (13) or plural (14), second person singular or plural (15), third person pronouns *he*, *she*, *it* and *they* (16)-(18) or third person noun phrases, singular and plural (19).

(13)   *I guess* that *I still the NORE[7] for this year.* (l, French, e-mail)
(14)   *We knew* that *X didn't speak German.* (o, German, e-mail)
(15)   *Do you also think* that *we should buy a Firewall program?* (c, Italian, e-mail)
(16)   *But he told us,* that *he can speak German!* (r, German, e-mail)
(17)   *it's about time* that *things get clear in this meeting story* (&, French, e-mail)
(18)   *they told me* that *they can give us some sample materials* (b, Italian, e-mail)
(19)   *Young teenagers (girls), think Ø their physical appearance is very important.* (f, French, e-mail)

The tense of the main verb was coded with a three way distinction: verbs in the present tense (20), verbs in the past tense (21), and sentences with no verb preceding the complementizer (22). Tagliamonte and Smith (2005, p. 304) had found that verbs in the present tense favoured the zero complementizer more than past tense verbs. There are relatively few tokens of sentences with no verb, so these will not be considered further for the factor of tense.

(20)   *I think also* that *X is in Geneva, isn't it?* (w, French, e-mail)
(21)   *In the meantime I found* that *he wrote his notes for fundraising on the IFMSA web page.* (c, Italian, e-mail)
(22)   *The fact* that *the computer is not always on minimizes greatly the possibilities for anyone to access it.* (j, French, e-mail)

In terms of additional elements in the verb phrase, the tokens were coded for whether there were no additional elements (23), whether the additional element was a modal verb (24) or whether the additional element was a negation form (25). The tokens which had both a modal and a negation form were given a separate code (26). Tagliamonte and Smith (2005, p. 304) found that 'simpler constructions' (i.e., those without additional elements in the verb phrase) favoured the zero complementizer and Torres Cacoullos and Walker (2009, p. 21) found the same.

---

[7] NORE = National Officer for Research Exchanges

(23) *I think Ø you'll understand why.* (c, Italian, e-mail)
(24) *To sum it up it <u>can</u> be said* that *SCOME-CH has to build a solid struc-ture for concrete projects.* (f, French, e-mail)
(25) *I <u>don't</u> think Ø it would be necessary to buy a multi-user license.* (j, French, e-mail)
(26) *I <u>cannot</u> promise* that *I can attend.* (*, French, e-mail)

Somewhat similarly to the previous factor, the tokens in the factor considering other modifiers in the verb phrase were coded for absence (27) or presence (28) of additional elements. Again, Tagliamonte and Smith (2005, p. 304) found that tokens without additional elements favoured the zero complementizer.

(27) *if you <u>Ø</u> thought* that *you had already won the portwine bottle* (p, German, e-mail)
(28) *I <u>really</u> hope* that *everyone arrived in Kopaonik as per travel-time-table!* (a, Italian, e-mail)

A number of specific lexical verbs were analyzed to determine how they affected the variability; any verb occurring frequently enough to allow it to be analyzed on its own was considered (*think, hope, tell, say* and *know*). The cut-off point for this was at least 24 tokens in the whole of the IFMSA data set. This is a far lower cut-off point than that used by Torres Cacoullos and Walker (2009, p. 16) in their analysis of spoken Canadian English. There, verbs with more than 200 occurrences were frequent and those between 10 and 49 occurrences were deemed to be very infrequent. Except for *hope,* which was not examined on its own by Tagliamonte and Smith (2005), the other specific lexical verbs are the same as the ones considered in Tagliamonte and Smith (2005) and similar to those found most frequently in Torres Cacoullos and Walker (2009).

Results

Overall Distribution

The nonnative speaker e-mail corpus yielded 576 tokens, and the native corpus 328. The breakdown of the tokens by native language and by *that* or zero complementizer is provided in Table 1.

Table 1 Overall distribution of the use of the zero complementizers (e-mails)

	% of zero	% of *that*	N
IFMSA overall	33	67	576
French	26	74	197
German	46	54	87
Italian	34	66	292
Natives	38	62	328

The occurrence of the zero complementizer form is far lower in all four groups than what was found in studies of native English speech. This is not entirely surprising, however, as it had been noted that speech has higher rates of the zero form than written data. The rates found in the present analysis are at a mid-point between the informal and formal texts that Elsness (1984) had considered. This further underlines how e-mails are a separate medium from both oral and written data and why it is crucial to use a native English control group of e-mailers to compare to the Swiss data.

Although it was initially hypothesized that some nonnative speakers would have much higher rates of *that* given their mother tongues do not have forms comparable to the English zero form, this does not seem to be the case exactly. The native group is not substantially different from the three nonnative groups in terms of percentage of zero complementizer forms, being at a mid-point between the German speakers and the French and Italian speakers, as can be seen from Figure 1.
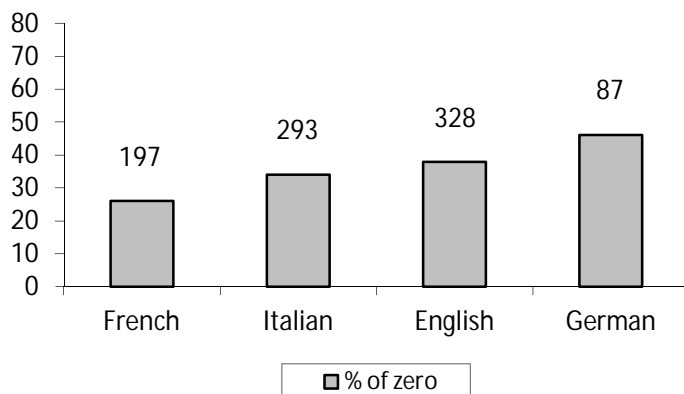

Figure 1 Percentage of zero complementizer (*N*s above bars)

A chi square test considering all four linguistic groups reveals that the differences between them are significant,[8] so, at this stage, we cannot state that the four groups are identical in their complementizer use. The three nonnative groups clearly make use of both variants.

Note that the German speakers use more of the zero complementizer than the native group; this may be due to interactions with the lexical verb. If the German speakers have a higher proportion of *think* than the other groups, for example, then this could explain the difference in zero complementizer use. We turn to this now.

Lexical Verb

Most studies found considerable differences in rates depending on the lexical verb (Kolbe's research only examined the most frequent lexical verbs in fact). This difference in rates was explained in two ways; first of all, some verbs, such as *think*, were more likely to have an epistemic meaning and more likely to use the zero form (Tagliamonte & Smith, 2005, p. 293), and secondly, the higher frequency verbs were also more likely to show high rates of zero complementizer (Torres Cacoullos & Walker, 2009, p. 15).

The present analysis examines *think, hope, tell, say* and *know* in detail, as these were the verbs which occurred most frequently in the nonnative data. The other verbs which were found to be more likely to use the zero complementizer did not occur frequently enough to warrant being considered individually.

Think. *Think* represents more than 20% of the overall tokens considered in this analysis, that is, 200 tokens across both corpora, making it the most frequently occurring verb, as was the case in previous studies. The percentages of *think* with a zero complementizer are lower for all four groups than what was found previously (Table 2). Nonetheless, the native speaker group is far closer to this near categorical average (with 90% zero) than the three nonnative groups (with ranges between 50 and 70%). Note that *think* has far higher rates of zero complementizer than the overall distribution in all four groups. Although the nonnative groups have lower rates than the natives, they share the direction of effect.[9] Despite the difference with the native group, the three nonnative groups show very similar rates; there is no significant difference between them.[10]

---

[8] $(df = 3, \chi^2 = 12.18, p < .01.)$

[9] A chi square calculation of the four groups confirms this is a significant result: $\chi^2 = 18.03, p < .001.$

[10] A chi square calculation of the three non-native groups is found to be not significant: $\chi^2 = 1.92, p < 1.$

Table 2 Distribution of complementizer forms for *think*

	% of zero	% of *that*	*N*s
IFMSA	52	48	161
French	51	49	51
German	68	32	22
Italian	54	46	80
Natives	90	10	39

The lower rates found in the nonnative speakers might also show that their e-mails are generally more formal than natives'; recall that for the relative pronouns, only the nonnative speakers had tokens of the highly formal variant *whom*. I will return to this point in the discussion.

Hope. *Hope* occurred frequently enough in the e-mail data for it to be considered individually, as there were nearly 100 tokens for the nonnative speakers and 25 in the native e-mails. Although most studies did not consider it separately from other verbs, Torres Cacoullos and Walker (2009, p. 16) found that the complementizer was deleted at a rate of 89%.

As in the case of *think, hope* demonstrates a high proportion of the zero complementizer in all four groups, with the German speakers being closest to the native percentages and the French speakers furthest away (Table 3).

Table 3 Distribution of complementizer forms for *hope*

	% of zero	% of *that*	*N*
IFMSA	62	38	95
French	56	44	27
German	81	19	16
Italian	60	40	52
Natives	88	12	25

Here again, although all the groups vary and show percentages of zero complementizer forms above 50%, there are differences between the groups. In this instance, the German speakers are very close to the native speakers, while the other two groups delete the complementizer about 20% less. The differences between the four groups are significant but, as for *think*, when only the three nonnative groups are considered, the difference is found not to be significant.[11]

Tell. The next high frequency lexeme is *tell.* Tagliamonte and Smith (2005. p. 301) had found rates of 64% percent of the zero form with *tell* in

---

[11] All four groups: $\chi^2 = 9.39$, *p* < .025. Three nonnative groups: $\chi^2 = 3.12$, *p* < 1.

their data, while Torres Cacoullos and Walker (2009, p. 16) had found 54%, both of which were considerably lower than the other verbs they examined and lower than the overall distribution (which was about 80%).

Note that directly reported speech has to be separated from indirectly reported speech with *tell*, because in direct speech there is no complementizer (Examples (29)-(31)).

(29)   *He told me 'I'm happy.'* – direct speech
(30)   *He told me* that/Ø *I was happy.* – indirect speech
(31)   *He told me* that/Ø *he was happy.* – indirect speech

While the number of tokens for *tell* is rather lower than for *think* and *hope,* the distributions of the various linguistic groups can still be analyzed, and once again there are considerable differences between the natives and the German speakers on one hand and the French and Italian speakers on the other (Table 4).[12]

Table 4 Distribution of complementizer forms for *tell*

	% of zero	% of *that*	N
IFMSA	23	77	31
French	11	89	9
German	44	56	9
Italian	15	85	13
Natives	45	55	11

Say. Tagliamonte and Smith (2005, p. 301) found that the zero complementizer occurred at a rate of 85% with *say* in their data, while Torres Cacoullos and Walker (2009, p. 16) found a rate of 73%. There are slightly over 40 tokens of it in the two e-mail corpora. The distribution of the tokens is not ideal as the French and native English speakers provide the majority of the tokens, so the results of the German and Italian speakers (with totals of three and four tokens respectively) cannot be considered to be truly indicative of the situation (Table 5), but they give us at least an idea of what is happening.

---

[12] The low number of tokens per cell means that it is not possible to establish whether these figures are statistically significant.

Table 5 Distribution of complementizer forms for *say*

	% of zero	% of *that*	*N*
IFMSA	21	79	24
French	18	82	17
German	33	67	3
Italian	25	75	4
Natives	37	63	19

As was the case for *hope* and *tell, say* occurs at a higher rate with the zero complementizer for the German and native English speakers. The French rate (18%) is considerably lower than the native group (37%). The native e-mail rate (37%) is much lower than what was found in speech.

Know. Similarly to *say*, Tagliamonte and Smith (2005, p. 301) found that *know* occurred with the zero complementizer at a rate of 85%, while Torres Cacoullos and Walker (2009, p. 16) find 66% deletion. There are around 50 tokens of *know* in the e-mail corpora, and in this case, it is the French and the German speakers that have a far lower number of tokens than the other two groups (Table 6). Insofar as it is possible to establish, the German and Italian speakers are similar to the native speaker rates.

Table 6 Distribution of complementizer forms for *know*

	% of zero	% of *that*	*N*
IFMSA	43	57	28
French	25	75	4
German	40	60	5
Italian	47	53	19
Natives	42	58	24

Table 7 presents a summary view of the percentages of zero comple-mentizer for all the verbs studied individually, while figure 2 presents these results in graph form. The rates of the different verbs alongside each other must be considered as this will allow us to establish whether the groups share the same patterns despite having different overall distributions. Figure 2 also plots the results from Tagliamonte and Smith (2005, p. 301) and Torres Cacoullos and Walker (2009, p. 16) to allow us to see how the oral data compares to the e-mail data.

Table 7 Percentage of zero complementizer by verb

	Overall	*Think*	*Hope*	*Tell*	*Know*	*Say*
English	38	90	92	45	42	37
French	29	51	56	11	25	18
German	49	68	81	44	40	33
Italian	32	54	60	15	47	25

What is most striking when examining Figure 2 is that the four e-mail groups show remarkably similar patterns; despite differences in percentages, the four e-mail groups have the highest rates for the same verbs. The hierarchy they all show is *hope > think* as the verbs with the highest rates of zero complementizer, with *tell, know* and *say* showing lower rates. The French and Italian groups have patterns that are marginally more similar to each other than to the other two groups and the same holds for the English and German groups; nevertheless the overall picture is that the three nonnative groups have similar patterns to the native control group.
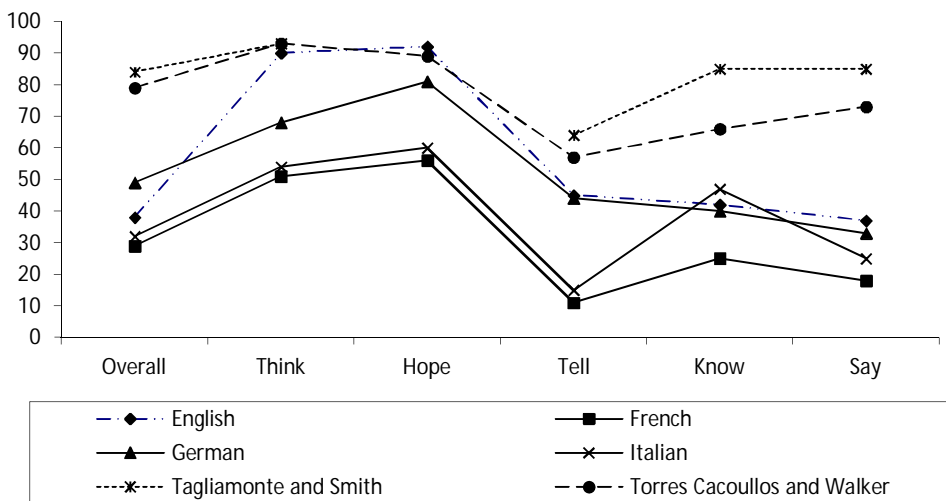

Figure 2 Percentage of zero by verb

Because Tagliamonte and Smith (2005) did not provide the rate of zero complementizer use with *hope* in their results, there is a gap in Figure 2. Although the rates of zero complementizer found by them and by Torres Cacoullos and Walker (2009) are considerably higher than for any of the e-mail groups, there are several points of similarity. *Tell* is the lexical verb with the lowest rate of zero complementizer in all six groups and *think* has one of the

highest rates. Similarly to the French and Italian groups, the results of studies examining speech show that *know* is used considerably more with the zero complementizer than with *tell.*

Overall, in terms of specific verbs, these results provide us with two major findings. First of all, the native e-mails show a distribution similar to the findings in previous studies in terms of specific verbs. The rates of zero complementizer are slightly lower than what was found in spoken data, but the near categoricity of the zero form with *think* is replicated here. The e-mail medium did not affect the variable rules underlying the distribution of the zero complementizer. Secondly, although the three nonnative groups have lower percentages than the native speaker control group, Figure 2 demonstrated that their patterning is very similar to the native speakers. Their percentages fall between those of the formal and informal text categories that Elsness (1984, p. 521) had examined, again underlying how difficult it can be to place e-mail in terms of written or oral registers.

Subject of Main Clause

Previous studies found that the subject of the main clause influenced complementizer choice in that first and second person subjects were more likely to occur with a zero form than third person forms. The data in the present study will be considered in terms of a four-way distinction with singular and plural subjects considered together; first person subjects, second person subjects, third person pronoun subjects and third person noun phrases. This follows the findings of previous studies.

The results of previous studies are matched by the nonnative e-mailers and the native e-mailers, as is demonstrated in Figure 3. First person subjects have the highest rates of the zero complementizer in all four groups and, except for the German speakers, these are followed by second person subjects.[13]

---

[13] Although Tagliamonte and Smith (2005) had found a considerable amount of interaction in terms of first person subjects and *think*, this analysis did not uncover the same categorical distribution of *I think* with the zero form for any of the linguistic groups.
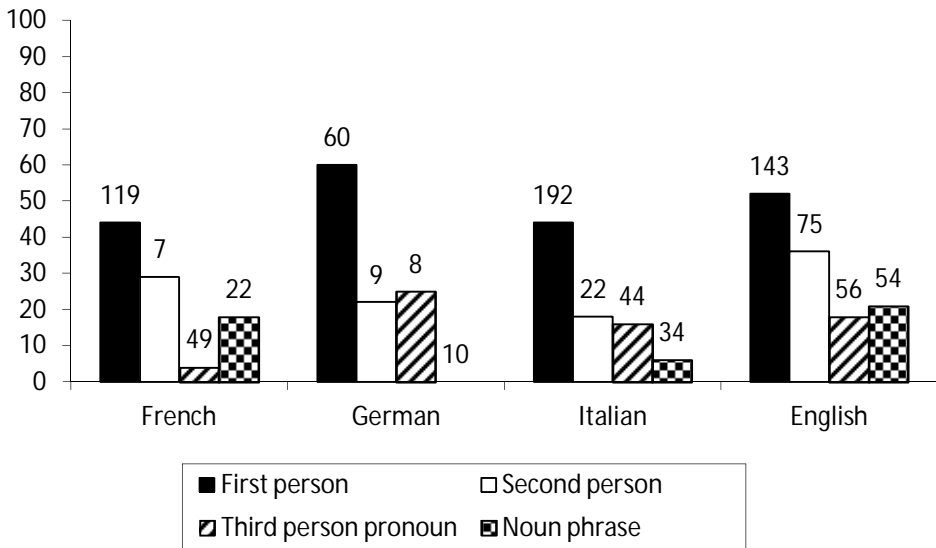
Figure 3 Percentage of zero by subject (*N*s above bars)

There are a number of differences in terms of hierarchy for the third person categories; however, the English and the French speakers have higher percentages of zero with noun phrases than with third person pronouns, while it is the opposite for the German and Italian speakers.

Overall, despite some differences in terms of third person subjects, the four groups are quite similar. The nonnative groups share the hierarchies of the native speakers and are varying their use of the zero complementizer according to the subject of the main clause in a very similar way to the native e-mailers.

Tense

Previous studies (Tagliamonte & Smith, 2005; Thompson & Mulac, 1991) found that verbs in the present tense are more likely to be used with the zero complementizer form than past tense verbs. The results for the present study are shown in Figure 4.[14]

Figure 4 demonstrates that, unlike the factors of lexical verb and the subject of the main clause, the groups are very different in terms of verb tense. Rather unexpectedly, the English group shows a higher proportion of zero deletion with past verbs than with present ones, which is at odds with

---

[14] Recall that instances with no verb are not considered here. This explains any discrepancies in total *N*s.

both the non-native e-mailers and with the findings for native speakers in previous studies. The difference between present and past complementizer use is not statistically significant for the English speakers however, so it is possible that it is not a true pattern and could merely be tied to the relatively low proportion of past tense forms opposed to the present tense tokens. It may also be because this analysis includes tokens of *I thought*, while Tagliamonte and Smith's (2005) study did not.
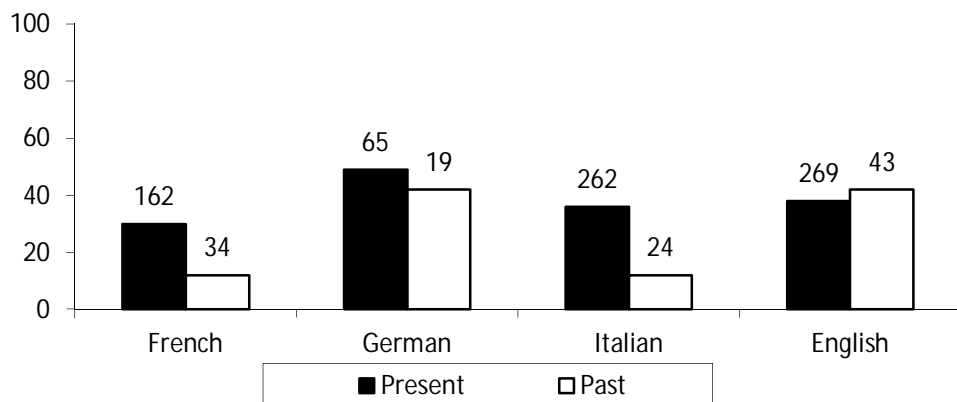
Figure 4 Percentage of zero by verb tense (*N*s above bars)

Additional Elements in Verb Phrase

Previous studies found that the presence of modal and negation forms lowered the likelihood of the zero complementizer being used. The results presented in Figure 5 consider this in terms of the e-mail corpora. Due to low figures, the presence of modal verbs and negation forms are combined.
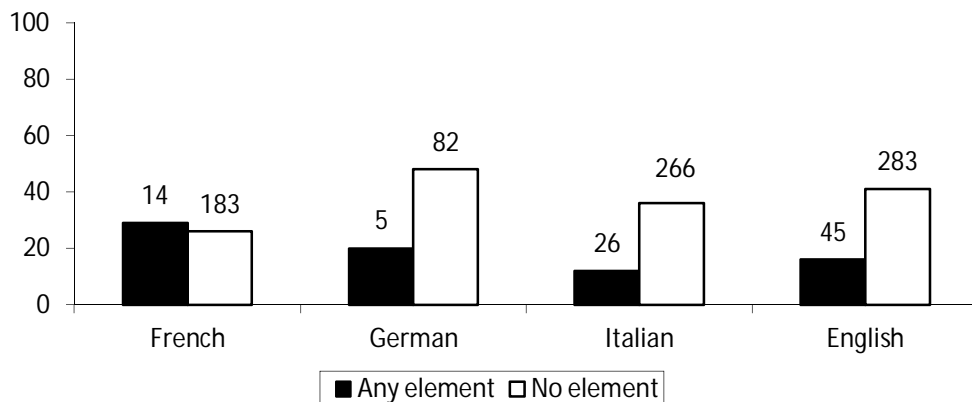
Figure 5 Percentage of zero deletion by elements in the verb phrase (*N*s above bar)

Although the hierarchy is very similar for the German, Italian and English groups, in that they all, as predicted, have higher rates of zero complementizer in clauses without additional elements, the overall low rate of tokens containing additional elements means that it is dangerous to attribute too much importance to these findings. The low number of tokens with any element in the verb phrase means that we cannot test whether the difference found in the French speakers is statistically significant.

Additional Elements in Main Clause

Tagliamonte and Smith (2005) and Torres Cacoullos and Walker (2009) found that additional elements in the main clause, such as adverbials, decreased the likelihood of a zero complementizer being used. Figure 6 analyzes the distribution in the e-mail data.
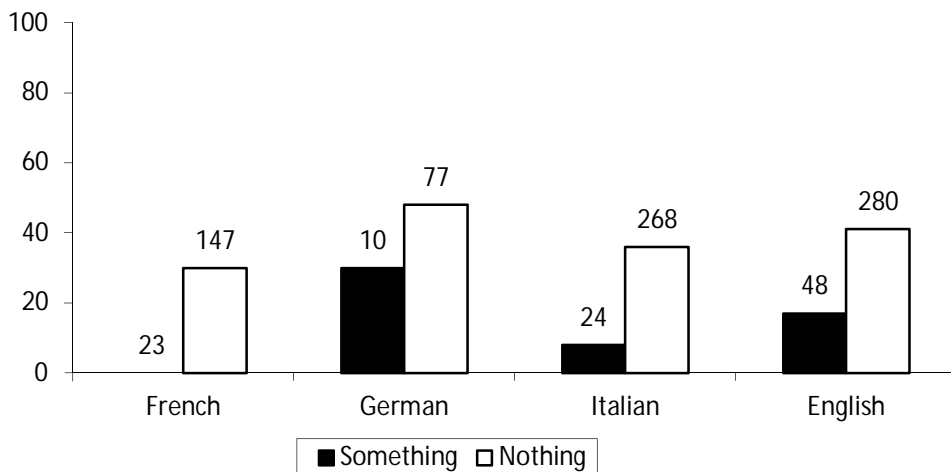

Figure 6 Percentage of zero complementizer with additional elements in the matrix clause (*N*s above bars)

All four e-mail groups show the predicted distribution; additional elements in the main clause lower the use of the zero complementizer form.[15]

---

[15] The data from each of the language groups was run individually in a multivariate analysis to fully establish which factors were likely to have a favourable effect on the zero complementizer, but as, by and large, they merely confirmed the patterns found in the overall distributions discussed here, I have decided not to include them (see Durham, 2007 for a full analysis).

Discussion

The various analyses have revealed similarities and differences among the natives and nonnatives, but also within the nonnative groups themselves. The points below summarize the main findings for each factor, before we turn to a full discussion of the results.

- Overall rates: The main difference in zero complementizer use is between the e-mail and the oral data. Whereas previous studies examining oral data had rates between 70-90% of zero, the e-mail groups, native and nonnative, are all around 35%. This is nonetheless higher than what Elseness (1984) had found for formal texts, but lower than the informal texts (1-15% and 52-58% respectively) underlining that the register of e-mails is more formal than oral data and informal written data. This affects native and nonnative speakers alike. The French and Italian groups are significantly lower than the English and German groups, however. Overall, the nonnative groups reach a close approximation of the native rates and clearly do use both variants.

- Lexical verb: Despite some differences in overall distribution of the zero form, the four e-mail groups have very similar patterns for the different verbs; *hope* and *think* are considerably higher than the other verbs. This matches the findings in Tagliamonte and Smith (2005) and Torres Cacoullos and Walker (2009). The nonnative groups have completely acquired this aspect of the variation.

- Subject of main clause: The four e-mail groups have the pattern found for oral data in Tagliamonte and Smith (2005); first person subjects have the highest rate of the zero variant followed by second person subjects and then other subjects. This factor constraining the variation in complementizers has been fully acquired by the nonnative speakers.

- Tense: The three nonnative groups show the expected favouring of the zero complementizer in present tense contexts. For the native speakers, on the other hand, there is no statistically significant difference between the two contexts, mostly likely due to low *N*s.

- Additional elements in the verb phrase: Here the order predicted was an increase of the zero complementizer in cases where there was no additional element in the verb phrase; all but the French speakers showed the expected order. This factor has been acquired by the German and Italian groups.

- Additional elements in the main clause: As for the previous factor group, the absence of additional elements was found to favour the

zero complementizer. This was found to be the case for all four e-mail groups, so they have acquired this aspect of variability.

Examining the whole range of factors influencing the use of complementizers, the overwhelming conclusion one can draw is that the nonnative groups are very similar to the native e-mailers. Although their overall rates are generally lower than in native data, the native constraints and patterns are there.

The nonnatives, for the most part, match native speaker patterns (either those of the e-mail control group or those considered in previous studies). *Think, tell, know* and *say* have higher rates of zero complementizer than other verbs, first person subjects are more likely to be used with a zero complementizer than other subjects, and elements, either in the verb phrase or the main clause, inhibit the use of zero. This demonstrates that the nonnative speakers have acquired many of the constraints which operate in English complementizer patterns.

The (Swiss) German speakers have a zero form used in a similar way to the English zero complementizer form in their native language and they are the nonnative group with the highest rates of zero complementizer. They are not the only group to show variability in the feature, however, as both the Italian and French groups also use the zero complementizer form. While the similarity between German and English might have benefited the German speakers in some ways, the fact that they and the other two nonnative groups share native speakers' hierarchies of constraints and ranges is due to more than surface similarity. Although we do not know what constraints operate on German complementizers, it is unlikely that they are the same constraints as in English. The German speakers (as well as the French and Italian speakers) are applying English variable rules for their use of the zero complementizer form.

Conclusion

This article has shown that despite the *that* and the zero complementizer forms not being explicitly taught to nonnative Swiss speakers of English, these speakers have nevertheless acquired the variable rules of native speakers. Not only do the patterns match the native e-mail corpus, but also what had been uncovered in previous analyses of English zero complementizers (Tagliamonte and Smith, 2005; Torres Cacoullos & Walker, 2009).

Despite there being no similar zero complementizer form in two of the three source languages, the French, German and Italian speakers have been able to integrate the variable rules of English complementizer distribution. This underlines the fact that there are a number of underlying syntactic distribution patterns which nonnative speakers can acquire without explicit teaching.

References

Bayley, R. (1995). Competing constraints on variation in the speech of adult Chinese learners of English. In R. Bayley & D. R. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 97-120). Amsterdam: John Benjamins.

Bayley, R. & Regan, V. (2004). Introduction: The acquisition of sociolinguistic competence. *Journal of Sociolinguistics, 8*(3), 323-338.

Dewaele, J.-M. (2004). Retention or omission of the *ne* in advanced French interlanguage: The variable effect of extralinguistic factors. *Journal of Sociolinguistics, 8*(3), 433-450.

Dewaele, J.-M. & Regan V. (2002). Maîtriser la norme sociolinguistique en interlangue française: Le case de l'omission variable de 'ne.' *Journal of French Language Studies, 12*, 123-148.

Durham, M. (2003). Language Choice on a Swiss mailing list. In B. Danet & S. Herring (Eds.),The multilingual Internet [Special issue]. *Journal of Computer Mediated Communication, 9*(1). Retrieved from: http://jcmc.indiana.edu/vol9/issue1/durham.html

Durham, M. (2007). *English in Switzerland: inherent variation in a non-native speech community* (Doctoral dissertation, University of Fribourg, Switzerland). Retrieved from http://ethesis.unifr.ch/theses/downloads.php?file=DurhamM.pdf

Dürmüller, U. (2001). The presence of English at Swiss universities. In U. Ammon (Ed.), *The effects of the dominance of English as a language of science on the non-English language communities* (pp. 389-403). Berlin: Mouton de Gruyter.

Dürmüller, U. (2002). English in Switzerland: From foreign language to lingua franca. In D. Allerton, P. Skandera, & C. Tschichold (Eds.), *Perspectives on English as a world language* (pp. 114-123). Basel: Schwabe.

Ellegard, A. (1978). *The syntactic structure of English texts: A computer-based study of four kinds of text in the Brown University Corpus*. Goteborg: Acta Universitatis Gothoburgensis.

Elsness, J. (1984). That or zero? A look at the choice of object clause connective in a corpus of American English. *English Studies, 65*, 519-533.

Herring, C. (2001). Computer mediated discourse. In D. Schiffrin, D. Tannen, & H. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 612-634). Oxford: Blackwell.

Hopp, H. (2007). Cross-linguistic differences at the syntax-discourse interface in off- and on-line L2 perfomance. In A. Belikova, L. Meroni, & M. Umeda (Eds.), *Proceedings of the 2nd conference on generative approaches to*

language acquisition North America (GALANA) (pp. 147-158). Somerville, MA: Cascadilla Proceedings Project.

Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269-293). Harmondsworth: Penguin.

Kachru, B. (Ed.) (1982). *The other tongue.* Pergamon Institute of English.

Kolbe, D. (2008). *Complement clauses in British English* (Unpublished doctoral dissertation). University of Trier.

Kucera H. & Francis, W. (1967). *Computational analysis of present-day American English.* Providence: Brown University Press.

Langman, J., & Bayley, R. (2002). The acquisition of verbal morphology by Chinese learners of Hungarian. *Language Variation and Change, 14*(1), 55-77.

Lyster, R. (1996). Question forms, conditionals, and second-person pronouns used by adolescent native speakers across two levels of formality in written and spoken French. *The Modern Language Journal, 80*, 165-182.

Mougeon, R., Nadasdi, T., & Rehner, K. (2010). *The sociolinguistic competence of immersion students.* Bristol: Multilingual Matters.

Mougeon, R. & Rehner, K. (2001). Acquisition of sociolinguistic variants by French immersion students: The case of restrictive expressions, and more. *Modern Language Journal, 85*, 398-414.

Mougeon, R., Rehner, K., & Nadasdi, T. (2004). The learning of spoken French variation by immersion students from Toronto, Canada. *Journal of Sociolinguistics, 8*(3), 408-432.

Nagy, N., Blondeau, H., & Auger, J. (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.

Penner, Z., & Bader, T. (1995). Issues in the syntax of subordination: A comparative study of the complementizer system in Germanic, Romance, and Semitic languages with special reference to Bernese Swiss German. In Z. Penner (Ed.), *Topics in Swiss German syntax* (pp. 73-290). Bern: Peter Lang.

Regan, V. (1995). Variation in French interlanguage: A longitudinal study of sociolinguistic competence. In R. Bayley & D. R. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 177-201). Amsterdam: Benjamins.

Regan, V., Howard, M., & Lemée, I. (2009). *The acquisition of sociolinguistic competence in a study abroad context.* Bristol: Multilingual Matters.

Rehner, K., Mougeon, R., & Nadasi, T. (2003). The learning of sociolinguistic variation by advanced FSL speakers: The case of *nous* versus *on* in immersion French. *Studies in Second Language Acquisition, 25*, 127-156.

Sankoff, G., Thibault, P., Nagy, N., Blondeau, H., Fonollosa, M., & Gagnon, L. (1997). Variation and the use of discourse markers in a language contact situation. *Language Variation and Change, 9*(2), 191-218.

Soars, L., & Soars, J. (1987). *Headway upper-intermediate.* Oxford: Oxford University Press.

Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research, 22*(3), 339-368.

Spencer, T. (1999). *The golden grammar book* (2nd ed.). Lausanne: Imprivite S.A.

Tagliamonte, S., & Smith, J. (2005). No momentary fancy! The zero complementizer in English dialects. *English Language and Linguistics, 9*(2), 289-309.

Thompson, S., & Mulac, A. (1991). The discourse conditions for the use of complementizer *that* in conversational English. *Journal of Pragmatics, 15*, 237-251.

Torres Cacoullos, R., & Walker, J. (2009). On the persistence of grammar in discourse formulas: A variationist study of *that. Linguistics, 47*(1), 1-43.

White, L. (2010). Second language acquisition at the interfaces. *Lingua, 124*(4), 577-590.