CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Variation and change in future temporal reference with *avoir* and *être**

Anna Tristram

Queen's University Belfast
Email: a.tristram@qub.ac.uk

ABSTRACT

Variation and change in the future temporal reference (FTR) sector in French has been the subject of numerous studies, from a variety of perspectives. Most studies consider the patterns of variation and evidence for change by looking at the verbal system as a whole. However, there are indications that some verbs differ significantly in their preference for one or other variant. *Avoir* and *être* are two such verbs. This study first examines the overall distribution of the inflected and periphrastic future with these two verbs in the ESLO corpus of spoken French, and considers the evidence for change. A multivariate analysis of the linguistic factors affecting variant selection in FTR with these two verbs reveals no exceptional effects; we thus explore other possible explanations for the exceptional distribution of FTR variants with these two verbs.

1. INTRODUCTION

The future temporal reference sector has long been of interest to scholars as an area in which competing forms exist to express the same meaning – i.e. to refer to an event which has not yet taken place. The main competing forms in modern French are the inflected future (IF), e.g. *je chanterai*, and the periphrastic future (PF), e.g. *je vais chanter*. The present tense is also used with future temporal reference (usually with a disambiguating temporal adverb), e.g. *demain je chante*. Many studies have explored the meaning of the forms, frequently attempting to uncover a unique meaning or function for each (e.g. Imbs, 1968; Vet, 1980; Wales, 1983, 2014; Confais, 1995; Abouda and Skrovec, 2015, *inter alia*). Others seek to examine the evidence for change in this sector (e.g. Blondeau, 2006; Poplack and Dion, 2009; Wagner and Sankoff, 2011; Poplack, Lealess and Dion,

CrossMark

2013, *inter alia*), with the periphrastic future argued to be in the process of replacing the inflected future (Söll, 1983; Blondeau, 2006; Poplack and Dion, 2009; Wagner and Sankoff, 2011).

Most studies tend to look at patterns for all verbs together. There are indications, however, that not all verbs behave in the same way: individual preferences for some verbs may be quite different from overall patterns. While it is possible to take this into account in statistical modelling, the extent to which some verbs may differ in their behaviour merits further attention. In particular, *avoir* and *être* are often mentioned as behaving quite differently from the majority of verbs (see e.g. Söll, 1983; Sundell, 1991; Poplack and Turpin, 1999 and discussion in Section 2.2). These two verbs, two of the most highly frequent in the French language, also exhibit exceptional behaviour in other domains of morphosyntax: for example, in relation to past tenses and the persistence of the *passé simple* (see below). This study investigates future temporal reference with *avoir* and *être* to establish how they differ from other verbs with respect to FTR, to attempt to understand what might lie behind this exceptionality, and to shed light on their behaviour in other domains of morphosyntax.

2. CONTEXT AND VARIANTS

The French inflected future, e.g. *je chanterai*, evolved from a Vulgar Latin periphrastic form consisting of an infinitive followed by an inflected form of *habere* ('to have'), e.g. *cantare habeo*. It can be found in some of the earliest texts of French (such as the *Serments de Strasbourg* of 842AD; see Comeau 2015: 345 for examples and discussion). This inflected future has co-existed for centuries alongside a periphrastic form constructed with *aller* + infinitive (e.g. *je vais chanter*). Fleischman (1982: 82) notes that the temporal meaning of PF dates from around the thirteenth or fourteenth century, becoming generalized in colloquial speech in the fifteenth century, and admitted to polite conversation and literary discourse during the sixteenth and seventeenth centuries.

In grammars and many scholarly studies of French, the approach has typically been to seek a unique meaning for each competing form – or to look for 'form-function symmetry' (Poplack and Dion, 2009). In grammars, a common and enduring argument has been that the periphrastic future is used for events which are temporally closer, and the inflected future for events that are more distant – hence the term *'futur proche'* or *'futur prochain'* (a term first used by the Abbé Antonini in his grammar of 1753). In fact, a huge range of meanings have been suggested in grammars for each form, with much overlap; a meaning proposed for the PF in one volume may be assigned to IF in another (see Poplack and Dion 2009 for a thorough survey of comments in grammars on FTR). In some key studies of FTR in French, the difference between IF and PF has been framed in terms of the *'visée prospective'* that the PF offers: while the IF is said to be detached from the moment of speaking, the PF retains a connection to the speaker's present (see, for example, Blanche-Benveniste et al., 1990; Fleischman, 1982; Jeanjean, 1988; also Abouda and Skrovec, 2017 for a recent fine-grained semantic analysis of these forms).

Work on FTR in the variationist paradigm takes a different approach. The competing forms are viewed as variants of the same underlying variable, and the realisation of one or other form in a given context is affected by a number of linguistic and extra-linguistic factors. D. Sankoff (1988: 153–156) argues that while it may be possible for a linguist to examine closely particular examples and detect some nuance of meaning which distinguishes them, in many contexts, these nuances are not relevant either in the intentions of the speaker, or in the interpretation of the listener. Thus, the forms are argued to be semantically equivalent in these contexts – i.e. differences in referential value are neutralized (see Poplack and Dion, 2009: 569).

Establishing equivalence of meaning is unproblematic at the level of phonology, since variants are not meaning-bearing. But where morphosyntactic (or lexical, or discourse-pragmatic) variants are concerned, meaning is necessarily involved, and the difficulty is in establishing that they are synonymous in every context. This issue gave rise to fierce debate in sociolinguistics from the late 1970s onwards, which will not be rehearsed here (cf. Blanche-Benveniste, 1977; Lavandera, 1978; Romaine, 1984; García, 1985; Cheshire, 1987; Winford, 1996; Gadet, 1997; see also Coveney, 2007). Since this time, the number of studies looking at variation 'above and beyond the level of phonology' (Sankoff, 1980) has grown significantly. It is now generally accepted that sociolinguistic variables can exist at any level of language, and that, with proper circumscribing of the variable context (see below), and careful examination of the variants in the contexts in which they occur, the study of them provides vital information about the ways in which variation at different levels of language is structured.

Tagliamonte (2012) notes that another way of establishing that variants are equivalents is to look at whether they are in complementary distribution: that is, where more of one is found, proportionally less of the other is found. This is demonstrably the case with the future variants (see below for details of quantitative studies). What is of interest, then, is to uncover the structured patterns of variation in usage of the competing forms, and to understand the factors that influence which variant is realised in a given context. Though variation does not always imply change, it is a necessary precondition for change to be possible (Weinreich, Labov and Herzog, 1968: 188). Hence, in this sector of long-standing variation, many studies have also been interested in establishing whether there is change taking place.

2.1. Previous studies

Quantitative studies of FTR in the French of France include Kahn (1954), which looks at spoken Parisian French and reports overall proportions of 72.3% IF and 27.3% PF. In another study of spoken Parisian French, François (1974) reports a slightly more elevated use of PF of 38.3% (with IF at 61.7%). Lorenz (1989) examines the same variety, reporting 57.1% IF and 42.9% PF.[1] Gougenheim

---

[1]Lorenz (1989) includes the futurate present (P) in the analysis, therefore these totals are recalculated to allow direct comparison with studies that do not include this form. See Villeneuve and Comeau (2016: 318) for a useful table summarizing previous studies.

et al. (1964) report 64.5% IF and 35.5% PF in a study which uses data obtained from speakers from a variety of locations (including a small number from outside France). Jeanjean's (1988) results concern the spoken French of Aix-en-Provence, where proportions of 57.8% IF and 42.2% PF are found. While these studies can give an idea of variation in the FTR sector in the French of France, and a sense of whether change might be happening over time (note the evolution of Parisian usage observable from 1954 [72.3% IF, Kahn] to 1974 [61.7% IF, François] to 1989 [57.1% IF, Lorenz]), a degree of caution is necessary when making comparisons. These studies take very different methodological approaches (in particular, it is not always clear whether modal uses of FTR variants are excluded), and the corpora used are also relatively small and unrepresentative (for example, only 101 tokens in total in Kahn, and 47 in François).

More recent work has addressed the question of change in FTR using larger oral corpora. Abouda and Skrovec's (2015) study uses data drawn from the ESLO (*Enquêtes Sociolinguistiques à Orléans*) corpus of spoken French (see also Abouda and Skrovec, 2017). They establish that PF has indeed become more common over the forty-year period examined (with IF accounting for 58% of uses in the earlier period examined, but only 28% in the later period), but that it is not a straightforward case of PF taking over all functions of IF. Rather, each form is argued to have certain temporal and modal functions. The authors argue that while PF is taking over some of the modal functions of IF, the latter retains some semantic 'niches'. For example, IF is favoured for uses termed *générique* by the authors, where the future '[...] présente le procès comme une prédication constante, caractérisant une classe d'individus' (Abouda and Skrovec, 2015: 15). The authors stress that in their account this category is separate from *typicalisation*, though elsewhere the label *générique* may be used for both of these (Damourette and Pichon, 1911–1939), while still other accounts call such uses *illustratif* (Bres and Labeau, 2013). While examples are given, the precise definition of these categories remains somewhat opaque, and it is also worth noting that in many variationist studies, such uses would be excluded, as they do not refer to future time (being categorized instead as gnomic or habitual uses; see Section 3.2).

The first variationist study of the French of France is Roberts (2012), based on the Beeching corpus of spoken French from the 1980s/1990s (Beeching, 2002; with data from a variety of locations, including Brittany, Paris, and southern France). With overall proportions of 58.8% PF and 41.2% IF, Roberts' study attests to the continued vitality of IF in this variety, despite the preference for PF overall (note the reversal of the proportions of PF/IF compared to the older studies described above). In another variationist study, Villeneuve and Comeau (2016) report on Vimeu French, a variety spoken in northern France (alongside Picard), with overall proportions of 62.2% PF and 37.8% IF.

Evidence from across the Atlantic strongly suggests that change is taking place in Canadian varieties. Synchronic studies of Canadian varieties spoken around the St Lawrence River, in Quebec and Ontario, have repeatedly shown that use of PF in spoken language far outweighs IF, to the extent that PF is now described as the default variant (Poplack and Turpin, 1999; Poplack, 2001; Poplack and Dion,

2009). For example, Emirkanian and Sankoff (1986) found proportions of 71% PF and 29% IF in Montréal.[2] Poplack and Turpin (1999) report similar proportions for Ottawa-Hull: 78.4% PF and 21.6% IF.[3] In Hawkesbury, Ontario, Grimm and Nadasdi (2011) found 86.5% PF and 13.5% IF based on data from 1978, rising to 89.5% PF and 10.5% IF in data from 2005 (Grimm, 2010).

Studies taking a diachronic perspective show that this preference is getting stronger over time. Poplack and Dion (2009) compare data from the Ottawa-Hull corpus and the nineteenth-century *Récits du français québécois d'autrefois* (Poplack and St-Amand 2007), and demonstrate that the proportion of PF used has increased from 61.3% in the nineteenth century to 78.4% in the twentieth century.[4] A note of caution is sounded by Wagner and Sankoff (2011), however. In their panel study of Montréal French, they uncovered an age-grading effect which seems to be slowing the pace of change in this variety: two-thirds of the speakers on the panel (n=60) actually *increased* their use of IF over time. Wagner and Sankoff view these results not as 'vitiating an apparent time interpretation, [but as] indicating that the rate of change may be slightly overestimated if age grading acts in a retrograde direction' (2011: 275). Blondeau (2006) reports similar results for a cohort of 12 Montréal speakers interviewed at three different times over a 25-year period.

In other Canadian varieties, however, the IF is still in widespread use. Indeed, Acadian varieties of French (those spoken in the Atlantic Provinces of New Brunswick, Newfoundland and Labrador, Prince Edward Island, and Nova Scotia) frequently show a strong preference for IF. For example, Comeau, King and LeBlanc (2016: 25) note that for the varieties spoken in Baie Sainte-Marie (Nova Scotia), L'Anse-à-Canards (Newfoundland), and the Iles de la Madeleine (Quebec), IF occurs at a rate of 38%, 24% and 39% respectively. King and Nadasdi (2003), examining usage in two communities on PEI, Abram-Village and Saint-Louis, and l'Anse-à-Canards, Newfoundland, found an overall preference for IF of 53%, leading them to conclude that claims of the decline of the inflected future 'in Canadian French *in general* seem premature' (2003: 332; emphasis in original).

2.2. Lexical differences

The studies discussed above present data on a range of verbs. A small number of studies discuss the distribution of future variants with individual verbs. In Söll's (1983) analysis of alternation of IF and PF in child language, the overall distribution of forms was 34% IF (n=155) and 66% PF (n=299).[5] These figures are reversed for *être*, with 67% IF (n=24) and 33% PF (n=12). Söll attributes this strong preference for IF to the extremely high frequency of *être* and its use

---

[2]Totals and proportions calculated by the author on the basis of data in Emirkanian and Sankoff's Table 1 (1986: 194).

[3]Poplack and Turpin (1999) is another of the few studies to include the Futurate Present (P) in the analysis. To allow direct comparison, as before, totals and percentages have been recalculated to remove tokens of P.

[4]These totals have also been recalculated to exclude tokens of P.

[5]Totals calculated by the author on the basis of figures given on p.19 of Söll, 1983.

as an auxiliary, two factors which, he argues, are related to its meaning. *Avoir* showed no strong preference either way in Söll's study – which begs the question, why not?, if an explanation for this pattern lies in high frequency and use as an auxiliary. Söll does not interrogate this further and it must be noted that the observation lacks explanatory power.

Sundell (1991) is an investigation of written French (the corpus consists of 50 contemporary novels), thus not directly comparable to many of the studies here. Nevertheless, it has some intriguing findings in relation to individual verbs: in particular, Sundell uncovers a strong preference for IF with both *être* and *avoir*. The overall distribution of variants in Sundell's study was 70% IF (n=4362) and 30% PF (n=1914). Sundell separates his tokens into three different types: 'Futurs non déterminés', 'Complément de temps' and 'Négation'. He focuses particularly on the effect of grammatical person on the distribution of forms. Overall, he observes a much higher frequency of IF for futures within the 'Complément de temps' and 'Négation' categories (85% and 83% IF respectively), but the figures for grammatical person follow more or less the same lines: Sundell observes that third person singular is generally found with a much higher rate of IF. This is particularly the case for *avoir* and *être*, where IF accounts for 90% and 82% of third person singular tokens respectively.

Sundell breaks this down further by examining the nature of the subject in tokens with third person singular. Impersonal *il* has the highest rate of IF use, 77% – while *on* has the highest rate of PF use, 50% (overall figures for third person singular are 68% IF and 32% PF). For impersonal *il*, Sundell finds that in fact PF is extremely rare with anything other than *falloir* ('il va falloir') and weather verbs. The three forms which account for the largest number of IF forms are *il y aura*, *il faudra* and *il sera* (accounting for almost 90% of IF forms with impersonal *il* between them). Sundell's data point to the existence of certain 'formulaic' uses, whereby one variant is vastly more likely than the other by virtue of the 'pre-formed' nature of these sequences (cf. Coveney, 1996, in relation to 'pre-formed sequences' in negation). Third person forms are among the most frequent, and are found with much higher rates of IF overall, and especially *avoir* and *être*. Certain forms account for very high proportions of these IF tokens: e.g. *il y aura*, *il faudra* and *il sera* in the case of IF with impersonal *il*.

Similarly, Poplack and Turpin observe that a small set of 'highly frequent and morphologically-irregular verbs' (1999: 155–156) including *vouloir, pouvoir, savoir, revenir, être, avoir* show an association with IF (albeit not very strong) in their data, and that these are the same verbs which are 'associated with the (otherwise non-productive) use of subjunctive morphology in the same corpus' (1999: 156). Roberts (2012: 105) also observes that 'a number of verbs appear to be selected almost categorically with the IF (e.g., *avoir* 'to have' 80%, n=31) and the PF (e.g., *travailler* 'to work': 88%, n=7)'.[6]

---

[6]In a study concerning peninsular Spanish, Blas Arroyo (2008) reports an effect for verb type and individual verb on alternation between IF and PF: while stative verbs in general preferred PF, *ser* ('to be') preferred IF (67%). The author notes that 'both the evolution of the verb *ser* throughout the history of Spanish (from its full meaning to its auxiliary role in a number of periphrases) and its

Table 1. ESLO corpus details

Sub-corpus	Interviews	Hours
1968 (ESLO1)	157	182.5
2008 (ESLO2)	150	150.0
Total	307	332.5

Though these studies note certain preferences for individual verbs, including a strong preference for IF with both *avoir* and *être* (perhaps more pronounced for the latter), in-depth analysis of what might underlie these patterns is lacking.[7] This study seeks to address this gap by focusing specifically on *avoir* and *être*, using data drawn from the ESLO corpus of spoken French. It seeks to address the following questions:

- Are the patterns of usage of the main FTR variants with *avoir* and *être* the same as those found in other studies which average data across all verbs?
- If not, how are the patterns different?
- Do the factors identified as significant in other studies have the same effect?
- If not, how is it different?
- What could explain any exceptional behaviour of these two verbs?
- Can this shed any light on the morphosyntactic behaviour of *avoir* and *être* in other domains?

3. METHODOLOGY

3.1. Corpus

The ESLO corpus of spoken French is composed of interviews recorded in Orléans in two periods: between 1968 and 1974 (ESLO1), and between 2008 and 2014 (ESLO2).[8] We henceforth refer to the 1968 corpus and 2008 corpus. The details of the two sub-corpora are presented in Table 1. The search interface allows the selection of different types of situation (interviews, family meals, shop interactions, debates, etc.); for the purposes of comparability, only interviews (*entretiens*) were selected for this study.

---

frequent appearance in formulaic phrases ('*lo que tenga que ser será*... "what is to be, will be...") could account for these differences in favor of MF [sc. morphological future, i.e. IF]' (Blas Arroyo, 2008: 111).

[7]Two studies which are also relevant here are Côté (2018) and Tremblay, Blondeau and Labeau (2019). Both of these studies consider the role of verb type; specifically, irregular vs. regular verbs, in FTR usage in L2 French and text messages (SMS) respectively. These two studies came out at a late stage in the development of the present study, but we discuss their relevant findings further below.

[8]<http://eslo.huma-num.fr/> [Last accessed November 2017] Corpus searches for this article were carried out in 2015.

3.2. Circumscribing the variable context

As in other variationist studies of FTR in French (e.g. Blondeau, 2006; Comeau, 2015; Poplack and Turpin, 1999; Poplack and Dion, 2009; Roberts, 2012, 2016; Villeneuve and Comeau, 2016; Wagner and Sankoff, 2011), this study considers variation in the forms used to express future meaning; it does not examine the totality of meanings that can be conveyed using morphologically future forms. Edmonds et al. (2017) advocate a concept-oriented approach to analysing FTR, in order to capture all verb forms used to refer to future time, arguing that such an approach avoids *a priori* assumptions based on morphological form. While such an approach is clearly valuable, in this study, to allow comparability with other variationist work on FTR, we focus on the main variants used to express future time, i.e. the inflected and the periphrastic future.

A small number of previous studies also include the Futurate Present (FP), but this form does not feature in the analysis here, for two main reasons. First, it is marginal in quantitative terms (other studies have found it accounts for less than 10% of tokens referring to future time; e.g. Poplack and Turpin 1999 – 7%; Poplack and Dion 2009 – 9%; though see Tremblay et al. 2019 for a discussion of FP in text messages). Second, it seems to be relatively stable over time. As Comeau and Villeneuve note, 'if there is ongoing linguistic change in how French expresses FTR, it primarily affects the IF and PF' (2016: 234). In addition, as noted by Blondeau (2006: 74), FP appears virtually categorically with an accompanying future adverbial (one of the factor groups considered here), which serves to disambiguate the usage of a present tense form referring to future time.

Searches of the ESLO (sub) corpora were conducted for all forms of *avoir* and *être* in the inflected future and periphrastic future (taking account of any material which may intervene between auxiliary and infinitive, e.g. *il va en avoir*). Since the number of tokens in the third person singular was far greater than for any other grammatical person/number, with a total of 1,190 tokens, the analysis concerns only data for this category.[9]

The first step in circumscribing the variable context is to eliminate any tokens of IF and PF which do not refer to future time. These exclusions fall into a number of categories, illustrated in the examples below: habitual actions (1); spatial movement (2); 'gnomic' utterances, i.e. timeless truths (3) (Fleischman, 1982: 132 following Greenberg, 1978, as cited in Wagner and Sankoff, 2011: 280; see also Imbs, 1960: 47); fixed expressions/sayings (4); hypothetical utterances (5).[10] Other exclusions included tokens where the interviewer primed the speaker; mis-

---

[9]Other studies have considered the effect of grammatical person on variation between PF and IF. The effect has typically been small, if detectable at all, and is thought to be related to formality/stylistic constraints rather than grammatical person *per se* (cf. Poplack and Turpin 1999; Roberts, 2012; Wagner and Sankoff, 2011). This factor is not considered here, though we acknowledge that the examination of third person tokens only means that caution must be exercised in the analysis regarding claims about wider applicability.

[10]Following Wagner and Sankoff (2011), contingent contexts (including those introduced by 'si', or where 'si' is implied) in which a probable future event hinges on some condition are included, but 'si' clauses which refer to a hypothetical situation without a specific temporal reference, are excluded. As Wagner and Sankoff note (2011: 279): 'Because the competition between IF and PF in the expression of the future was our

assigned tokens (error in transcription); fragments, repetitions and self-corrections; quoted speech; otherwise unclear tokens which could not be reliably categorized. After these exclusions, 649 tokens remained.

(1) il y a des jours où on travaillera beaucoup plus euh que d'habitude et puis il y a certains jours où euh ce **sera** très très calme très relax (1968-MK532)
(2) on fera construire par là mais là on **va être** ici un bon moment (1968-PY94)
(3) euh disons que dans une maison de jeunes où y a beaucoup de fils d'ouvriers y **aura** peu de fils de de bourgeois (1968-RF211)
(4) on dit toujours ça **sera** toujours Paris hein (2008-FJ30)
(5) parce que si on compte sur les élus sur la mairie sur les institutions y **aura** rien (2008-MF363)

A second step is to exclude any invariant contexts (i.e. which are found with 0% or 100% of one or other variant, and thus no variation). In many studies of FTR in Laurentian French,[11] negative contexts have been found to be categorical, in that only IF is selected in such contexts. In the present study, both IF and PF were found in negative contexts – i.e. there was no categorical effect – and negative contexts are therefore included in the analysis. Certain phrases with *avoir*, however, were revealed to be categorical (no such phrases were identified for *être*). Utterances which contained the phrase *avoir besoin* ('to need'; n=16) occurred only with IF, in both sub-corpora. Utterances with the phrase *avoir …ans* ('to be … years old'; n=25) occurred only with PF, with one exception. For other phrases (e.g. *avoir du mal*, 'to have trouble', *avoir envie de* 'to want/wish to', *avoir lieu* 'to happen'), the number of tokens was too low to establish categoricity. Tokens were also coded for whether they occurred within the phrase *il y a* ('there is/there are'), in its various permutations (e.g. *il y aura*, *il va y avoir*, with pronouns *y* or *en* etc.). No categorical pattern was discernible here, and behaviour with *il y a* did not seem exceptional (i.e. proportions of IF/PF were similar to those in corpus as a whole).[12] Contexts which did not admit variation (total n=41) were excluded from the subsequent analysis, leaving 608 tokens.

3.3. Coding

The next step is to code the data for factors of potential interest. One of the aims here was to establish whether any of the factors commonly investigated in connection with the topic have an unusual effect on distributional patterns with *avoir* and *être*. Factors identified as of interest in previous literature were therefore selected for testing here (see references below for each factor; findings vary across studies as to whether the factor was significant, but if it was

---

principal concern, tokens were excluded from the dataset if they did not unambiguously refer to future time, but instead represented instances of modal or aspectual usage of futurate forms'.

[11] The term 'Laurentian French' refers to varieties spoken in Quebec, Ontario and in western Canada.

[12] This was confirmed in a chi-square test, where the difference between patterns with *il y a* and where there was no particular phrase employed was not significant in either sub-corpus.

identified as being of potential interest, it was included here). Can any differences in the effect of factors on these two verbs help to explain their exceptionality? The factors investigated and brief details of coding procedure are as follows:

Linguistic factors

Corpus (i.e. date of recording): tokens were coded as either 1968 or 2008. (See e.g. Poplack and Dion, 2009).

Verb: *avoir* or *être*. (See e.g. Söll, 1983; Sundell, 1991; Poplack and Turpin, 1999; Roberts, 2012).

Proximity: Coding for this factor was into five categories, for events occurring within the hour, day, week, longer than a week, or which were 'continual' – i.e. existed before the time of the utterance and would continue to exist in the future envisaged in the utterance. To avoid circularity and subjectivity, coding was done on the basis of adverbs or other time-indicating phrases in the immediate or wider context. (See e.g. Poplack and Turpin, 1999; Poplack and Dion, 2009; Grimm and Nadasdi, 2011; Roberts, 2012).

Sentential polarity: Tokens were coded as positive or negative. (See e.g. Emirkanian and Sankoff, 1986; Poplack and Turpin, 1999; Poplack and Dion, 2009; Wagner and Sankoff, 2011; Roberts, 2012, *inter alia*). Presence/absence of *ne* and the negative item (e.g. *pas, plus, jamais* etc.) were also noted.

Contingency: Events were coded as either 'contingent', where the future eventuality was dependent on some other condition being fulfilled, or 'assumed', where there was no such condition and the event was assumed to be valid. This is the approach used in Poplack and Turpin (1999: 153), which itself is based on Fleischman (1982). (See e.g. Poplack and Turpin, 1999; Grimm and Nadasdi, 2011).

Adverbial modification: Tokens were coded for adverbial modification in three categories: absence of adverbial specification; presence of non-specific adverbial; and presence of specific adverbial, for example '*lundi prochain*' or '*en fevrier*'. (See e.g. Emirkanian and Sankoff, 1986; Poplack and Turpin, 1999; Poplack and Dion, 2009; Grimm and Nadasdi, 2011).

Presence of *quand* 'when': Tokens in the present study were coded for presence or absence of *quand*. (See e.g. Emirkanian and Sankoff, 1986; Grimm and Nadasdi, 2011).

Presence of certain phrases: During initial inspection of the data, it was noted that a number of phrases occurred quite frequently, e.g. *avoir … ans* ('to be … years old') and *avoir besoin* ('to need'). These and other regularly-occurring phrases were coded for, as there seemed to be tendencies for a particular phrase to appear with either IF or PF exclusively.

Sociolinguistic factors

Age: This was initially coded into seven categories (10–15 years old, 15–25, 25–35, 35–45, 45–55, 55–65, 65+, plus 'Not stated'). These were subsequently collapsed into three main groups (15–35, 35–55, 55+) to correspond to major life stages of young adulthood, middle age, and older/retirement age, as in some cases an age

band was empty or contained few speakers.[13] (See e.g. Emirkanian and Sankoff, 1986; Poplack and Turpin, 1999).

Gender: This was coded as either male or female, following the labels used in the original corpus. (See e.g. Villeneuve and Comeau, 2016).

Socio-economic status: The categorization schema for SES in the ESLO corpus (useful for fine-graded distinctions but less of a concern here) was simplified into two categories, Working Class (WC) and Middle Class (MC), on the basis of the INSEE *professions et categories socioprofessionnelles* categorization schema.[14] (See e.g. Emirkanian and Sankoff, 1986).

Education level: The ESLO education levels were also simplified into a three-way distinction: Bac- (no qualifications beyond the end of compulsory schooling), Bac (*baccalauréat*), and Bac+ (university-level qualifications). (See e.g. Roberts, 2012; Villeneuve and Comeau, 2016).

In the next section, we first present a distributional analysis of the two variants in the corpus to examine the general tendencies and effect of the different factors, with chi-square tests for significance where appropriate. Second, we present the results of a multiple logistic regression which examines the relative weight of the different factors. Factors which were not significant or for which insufficient numbers of tokens were obtained are not discussed any further in the quantitative analysis. These were: presence of *quand* (insufficient tokens); contingency, age, gender, education level and socio-economic status (not significant). Social factors are discussed briefly in a qualitative analysis of tokens of PF found in negative contexts (end of section 4.3).

4. RESULTS

4.1. Overall

The overall distribution of IF and PF in the whole dataset reveals a marked preference for IF with *avoir* and *être* (69.1%; Table 2), in contrast to the findings of most previous studies looking at all verbs together, in which a preference for PF is found (with the exception of some Acadian varieties, as discussed above). The proportions seen here more closely resemble those reported in the older quantitative studies of spoken French discussed above (e.g. Kahn 1954; François 1974; Lorenz 1989; Gougenheim et al. 1964) and those on written French (e.g. Sundell 1991 as discussed above; see also Lesage and Gagnon 1992). Thus, the behaviour of these two verbs looks highly conservative when compared to overall patterns of IF/PF usage with all verbs.[15]

---

[13]In the analysis, speakers in the 'Not stated' age category were left out for reasons of low token numbers.

[14]<http://www.insee.fr/fr/methodes/default.asp?page=nomenclatures/pcs.htm> [Accessed August 2016]

[15]It is possible that the particularities of the corpus exert an effect here. As noted above, only '*entretiens*' (interviews) were selected from the possible recording situations in the ESLO corpus, to allow comparability with other variationist studies of FTR, which are generally based on corpora of speech data drawn from semi-structured interviews. It is widely acknowledged in the sociolinguistic literature that such speech data have their limitations, including the extent to which they can be said truly to represent the vernacular. Nonetheless, given the principled construction of the ESLO corpus, it was judged unlikely that the particularities of the corpus are solely responsible for the high rate of IF usage observed here, and the analysis was carried out to investigate further. I am grateful to an anonymous reviewer for raising this point.

Table 2. Overall distribution of future variants

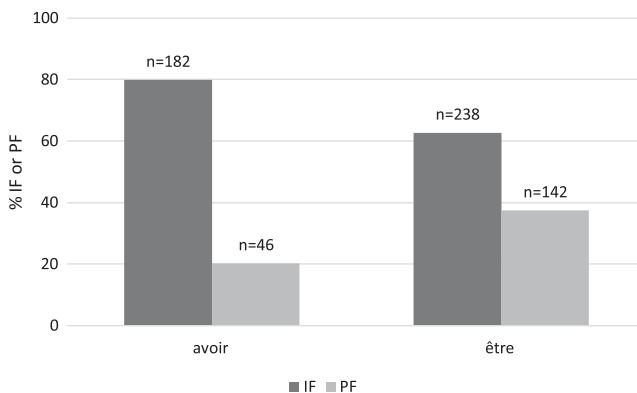|	N	%
IF	420	69.1
PF	188	30.9
Total	608	

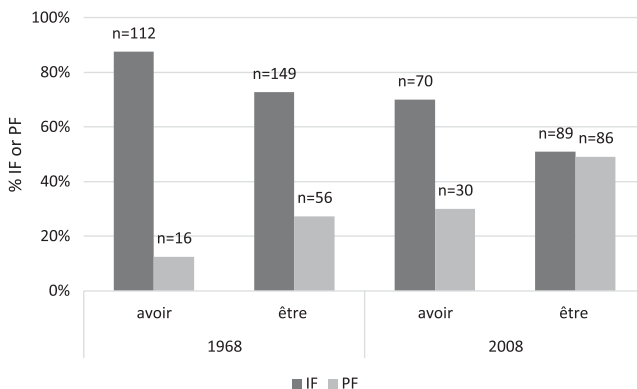**Figure 1.** Distribution of variants by verb.

Figure 2. Distribution of future variants by verb and corpus.

When the overall data are split by verb (Figure 1), it is evident that *avoir* has a much stronger preference for IF than *être*, with almost 80% IF and 20% PF compared to around 60% IF and 40 PF% for *être*.

4.2. Change

Given the conservative behaviour of *avoir* and *être* overall, it is all the more interesting to observe that there is nonetheless evidence of diachronic change taking place: the proportion of PF has increased in the later period, at the expense of IF. Figure 2 shows that the two verbs behave similarly in the earlier period, with both strongly favouring IF: *avoir* 87.5%, *être* 72.7%. In the later period, this effect has weakened, especially for *être*, for which an almost 50/50 distribution can be observed. *Avoir*, meanwhile, still strongly prefers IF (70.0%). These results were significant in a chi-square test (*avoir*: $\chi^2=10.676$, df$=1$, p$=0.001$; *être*: $\chi^2=19.215$; df$=1$; p<0.0001).

4.3. Multivariate analyses

A mixed-effects multiple logistic regression analysis was carried out on these data using Rbrul (Johnson, 2009).[16] An advantage of this type of analysis is that it allows all factor groups to be included simultaneously in one statistical model and can give an indication of their relative strength. Mixed-effects (as opposed to fixed-effects) models can also take into account potential random effects such as speaker and word-level variation, and factors are only selected as statistically significant when their effect rises above inter-speaker variation (Johnson, 2009: 365). The results of such an analysis provide 'three lines of evidence' (Tagliamonte, 2012: 122; Poplack and Tagliamonte, 2001: 92; Tagliamonte, 2002: 731) with which to understand and explain the variation in question. These are: (i) statistical significance of the effect (at the p$=0.05$ level); (ii) magnitude of the effect, evident from the range between the highest and lowest factor weight in a factor group; and (iii) direction of effect, shown by the hierarchy of factor weights within a factor group (Poplack and Dion, 2009: 572).

A separate analysis was carried out for each verb and the results are shown in Table 3 (*avoir*) and Table 4 (*être*). The tables show the factor groups which were significant for each verb. IF is selected as the application value in order to allow comparability with previous studies. Regression coefficients are expressed as both a log-odd and a factor weight (weighted probability). A positive log-odds value indicates that the factor favours the application value; a negative log-odds value indicates a disfavouring effect, while a value of 0 is neutral. The log-odds value gives an indication of the magnitude of the effect. A factor weight greater than 0.5 indicates that the factor favours the application value variant (i.e. IF), while a factor weight of less than 0.5 indicates that it disfavours it.

The multivariate analyses indicate that of all the linguistic factors examined, only three had a significant effect in selection of future variants with *avoir*: sentential polarity (p$=0.00157$), corpus (i.e. date of recording; p$=0.00208$) and proximity (p$=0.0377$). For *être*, these three factors were also significant (proximity p$=0.0000129$; corpus p$=0.0000248$; polarity p$=0.0298$), as well as adverbial modification (p$=0.0279$). Of the social factors examined, none was significant

---

[16]A new version of Rbrul with a graphical interface ('Shinybrul') was released in May 2016. The analysis presented here was carried out in the previous version ('Oldbrul'). See <http://www.danielezrajohnson.com/rbrul.html> [last accessed September 2018].

Table 3. Multivariate analysis results for *avoir*

Factor group	Levels	Log odds	Factor weight	IF/IF+PF	n
Proximity	> week	0.844	0.699	0.822	191
	day	0.744	0.684	0.667	3
	continual	0.413	0.602	0.826	23
	hour	−0.550	0.366	0.500	2
	week	−1.481	0.185	0.333	9
	Range	*51*			
Sentential polarity	negative	0.989	0.729	0.953	43
	positive	−0.989	0.271	0.762	185
	Range	*46*			
Corpus	1968	0.595	0.644	0.875	128
	2008	−0.595	0.356	0.700	100
	Range	*27*			

Application value = IF
n = 228, df = 8, log likelihood = −104.199, overall proportion = 0.798, centred input probability = 0.897
Factors not selected as significant: Contingency; Adverbial modification; Age; Sex; SES; Education

Table 4. Multivariate analysis results for *être*

Factor group	Levels	Log odds	Factor weight	IF/IF+PF	n
Proximity	continual	1.491	0.816	0.757	37
	> week	1.281	0.783	0.647	309
	day	0.129	0.532	0.600	10
	week	−1.444	0.191	0.182	11
	hour	−1.456	0.189	0.154	13
	Range	*63*			
Adverbial	specific	0.861	0.703	0.667	24
modification	non-specific	−0.231	0.443	0.699	103
	absent	−0.630	0.347	0.593	253
	Range	*35*			
Sentential	negative	0.533	0.630	0.800	30
polarity	positive	−0.533	0.370	0.611	350
	Range	*26*			
Corpus	1968	0.528	0.629	0.727	205
	2008	−0.528	0.371	0.509	175
	Range	*26*			

Application value = IF
n = 380, df = 10, log likelihood = −223.028, overall proportion = 0.626, centred input probability = 0.58
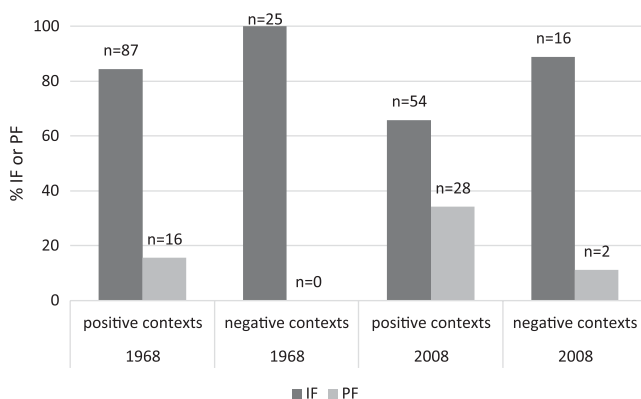Factors not selected as significant: Contingency; Age; Sex; SES; Education

Figure 3. Distribution of variants with avoir in positive and negative contexts in the two corpora.

for either verb – indicating a lack of social conditioning for this variable, which has also been the case in many previous studies (see for example those cited above), though a larger sample would clarify this result.

The results for 'corpus' confirm what we have already seen above: that these two verbs are participating in the more general change that is taking place, with PF becoming more common for both verbs (though for *avoir*, in the 2008 corpus, IF is still the majority variant).

The results for proximity indicate that the effect of this factor group was broadly similar for both verbs: for events longer than a week away and those which are 'continual', IF is preferred. Events which are set to occur within the day also prefer IF, though this effect is weaker, and in the case of *être*, very slight indeed (i.e. factor weight very close to 0.5). For both verbs, events within an hour and within a week disfavour IF – i.e. are more likely to be found with PF. However, due to the low token numbers in the hour, day and week categories (most acute for *avoir*), caution must be exercised when interpreting the ordering of the levels within this factor group. We can perhaps be most confident about the result for the 'longer than week' category, since this has the largest number of tokens. This factor group has the largest range for both verbs, and for *être*, the lowest p-value (though sample size can be a crucial factor when calculating p-values, therefore we will use them here with caution). These results are therefore broadly in line with previous studies where proximity was found to be a significant factor: events which are further away in time favour IF.

Adverbial modification was significant only for *être*. Contexts where there was a specific adverbial favoured IF; contexts where there was a non-specific adverb, or no adverbial modification, favoured PF, though for non-specific contexts this effect was very weak. This is broadly in line with previous studies, where the presence of an adverb (specific or non-specific) has been found to favour IF. It is somewhat difficult to explain why this factor should be significant for *être* only; it could possibly be an artefact of sample size (the number of tokens with *être* being larger).

Polarity was a significant factor for both verbs; more so for *avoir*, with a larger range and lower p-value. In negative contexts, IF was strongly favoured; in positive

Table 5. Negative tokens with PF

Negative tokens with PF	1968	2008	Total
Avoir	0	2	2
Être	2	4	6
Total	2	6	8
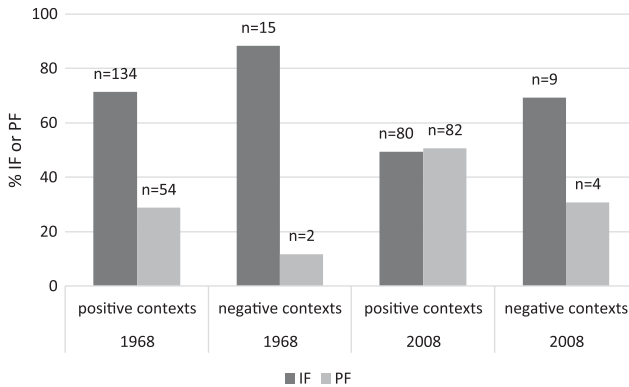
Figure 4. Distribution of variants with être in positive and negative contexts in the two corpora.

contexts, IF was less strongly favoured, though it is still the majority variant for both verbs (Figure 3 and Figure 4). Nevertheless, we do find some examples of PF in negative contexts (see below).

Given that in some previous studies PF is entirely absent from negative contexts, it is interesting to examine more closely the tokens of PF that we find in negative contexts in this study. There are eight tokens of PF which occur in negative contexts. The majority of these are with *être*, as shown in Table 5 and examples (6) to (13).

Tokens of PF in negative contexts with être

(6) alors le CES c'est un enseignement court qui se qui conduit l'enfant à arrêter ses études ou entendez non c'est le CEG c'est je me suis trompée voulez-vous faire une correction alors je parle des C bon oui ça va ça **va pas être** très brillant le CEG euh conduit l'enfant euh à un enseignement court qui s'arrêtera à la fin de la scolarité (1968-IG298)

(7) vous savez ça ça ça ça vient de soi hein c'est toujours pareil alors moi j'ai jamais lu de livre comme ça ça peut être utile comme ça **va pas être** utile ça" (1968-MH539)

(8) ça ça **va pas être** possible hein (2008-BI58)

(9) si on va sur un autre chantier on va être de trop on va se gêner ça **va pas être** utile (2008-BV1)

Table 6. Profiles of speakers who produced a negative token with PF

Speaker ID	Corpus	Age	Sex	Education	SES
MX953FEM	2008	15to35	F	Not stated	Not working
BV1	2008	15to35	M	Bac+	WC
MH539	1968	35to55	M	Not stated	WC
JR18	2008	35to55	F	Bac	MC
IG298	1968	35to55	F	Bac-	MC
BI58	2008	35to55	F	Bac+	MC
UZ57	2008	35to55	F	Bac-	MC
HV753	2008	55+	M	Bac	MC

(10) alors le Tibet on va laisser tomber parce que je suis objective euh ça **va pas être** euh demain qu'on va pouvoir remettre les pieds au Tibet facilement (2008-JR18)

(11) [il] va continuer à faire des petits articles en tant que pigiste mais il **ne va plus être** euh rédacteur en chef (2008-UZ57)

Tokens of PF in negative contexts with avoir

(12) il faut faire un même procédé parce qu'on a une usine d'entretien qui est à la Source on **va pas avoir** un deuxième (2008-HV753)

(13) je j'ai su y a deux jours que en fait le transport euh on on **va pas avoir** de transports pendant les deux mois de vacances (2008-MX953FEM)

Following Poplack and Dion (2009), we could view the tokens of PF in negative contexts as the residue of an incomplete change involving the embedding of IF as the only acceptable form in these contexts. However, this interpretation is not strongly supported by the evidence available for the French of France (see above), and especially since there are actually more tokens of PF in negative contexts in the later corpus.

A qualitative analysis of the linguistic context of these examples, in an attempt to characterize the conditioning of such examples (cf. Poplack and Dion 2009: 575, fn13), does not reveal any strong patterns, except perhaps that most are found with an impersonal subject, either 'ça' (with *être*) or 'on' (with *avoir*) (cf. discussion of Sundell, 1991 above). The only personal subject is *il* in (11). The following context features a variety of different word types: mostly adjective, but also noun, adverb. Several are preceded by periphrastic futures with other verbs (e.g. (9), (10), (11)) which could prime for PF in the tokens examined. The negative item in the majority of cases is *pas*, and *ne* is usually not present (presence of *ne* has been shown in other studies to inhibit PF; Roberts, 2012). One token, (11), stands out as different from the others. This token has personal

Table 7. Distribution of variants by sentential polarity, three-way analysis

Sentential polarity, 3-way analysis	IF		PF		
	n	%	n	%	Total
Positive contexts	355	66.4	180	33.6	535
Negative contexts, *ne* absent	49	87.5	7	12.5	56
Negative contexts, *ne* present	16	94.1	1	5.9	17
	420	69.1	188	30.9	608

il as the subject, *ne* is present, and the negative item is '*plus*'. Despite these rather exceptional features, some of which mark a more formal context (IF having been associated with formal contexts – see e.g. Roberts, 2012), PF is still selected.

As far as the social characteristics of the speakers are concerned, amongst the speakers who produced a token with PF are speakers from all age groups, both sexes, all SES groups and all education levels (Table 6). They all produce IFs elsewhere (with *avoir*, *être* or other verbs), therefore are not categorical PF users (cf. Wagner and Sankoff, 2011: 303). Thus it seems that these examples cannot be explained away as somehow 'exceptional' (cf. Wagner and Sankoff, 2011 – the only negative tokens with PF found in their study were hesitations or reformulations), and must be viewed as part of the normal workings of the variable grammar underlying alternation between IF and PF in this variety of French.

Following Roberts (2012: 102), a further analysis was conducted on the results for sentential polarity. Table 7 confirms Roberts' findings, i.e. that the incidence of IF increases as formality increases (the presence of *ne* being a well-known marker of formality in spoken Metropolitan French; Ashby, 1981; Coveney, 1996; Armstrong, 2001).

5. DISCUSSION

The results here show that these two verbs have a much stronger preference for IF than most other verbs, when compared to the results of previous studies, which have looked at patterns for all verbs together.[17] However, the effect of the factors involved in determining the distribution of IF and PF is much the same here as has been found in other studies. In the light of this, what can explain the exceptionally high rate of IF use with *avoir* and *être*?

These two verbs are highly frequent and morphologically irregular.[18] We know from Bybee's work (Bybee and Thompson, 1997; Bybee, 2003) that while for

---

[17]In order to fully determine the extent to which *avoir* and *être* are exceptional, one would need to examine all verbs individually, which is beyond the scope of this study. However, lexical differences can be taken into account in mixed-effects statistical modelling, by including 'lexical item' or 'word' as a random factor. This means that the model only reports as statistically significant factors whose effect rises above the effect of individual, word-level (or, in the case of individual speakers, speaker-level) differences (cf. Johnson, 2009).

[18]See frequency calculations in Appendix.

phonology, high frequency promotes (reductive) change, it may – paradoxically – help to preserve conservative morphosyntax.[19] Bybee (2003: 621) notes that 'repetition affects morphosyntax by ensuring the retention of older characteristics'. In this case, *avoir* and *être* favour an older morphosyntactic variant, IF, over a newer one, PF.

Other studies of future temporal reference which have noted this tendency include Blondeau and Labeau (2016), who examine FTR in French television weather bulletins, and found that the only significant linguistic factor constraining variation in their data was 'type of verb'. Irregular verbs (which tend to be the most highly frequent) favoured IF, while regular verbs disfavoured it. The authors note that irregular forms tend to be more conservative, but do not pursue this further. They do however note that studies of other sociolinguistic variables have found similar tendencies: Poplack (2001) in relation to the subjunctive, and Labeau (2015) in relation to the *passé simple*. In other areas of morphosyntactic variation, there are similar trends: Tristram (2014), examining the case of verbal agreement with collective nouns in French in both oral and written French, shows that more conservative patterns are observed with the most frequent items (e.g. *majorité*). Lindqvist's (1979) study of the imperfect subjunctive in written French observes that the choice between present or imperfect subjunctive after a main clause past tense verb is conditioned by a number of factors, one of which is the identity of the verb taking the subjunctive. When this is *avoir* or *être*, the conservative form, imperfect subjunctive, occurs at a rate of 65%, dropping to 48% for all other verbs.

But an explanation along the lines of mere frequency does not suffice. It is helpful to look to other areas of morphosyntactic variation for insights. Waugh and Monville-Burston's work on past tenses in data from newspaper usage, while not directly comparable as it concerns written language, nonetheless provides support for the hypothesis that highly frequent items tend to associate with older morphosyntactic forms. In relation to the competition between the newer compound past (*passé composé*) and older, simple past (*passé simple*), they note that:

> The result of the competition between SP [sc. simple past] and CP [sc. compound past] has been [...] that the newer CP has expanded its terrain, taking on the more general meaning (present and past perfect). Correspondingly, the older SP has become more specialized: it is lower in frequency, restricted to specific kinds of texts, associated with certain verbs of high frequency (such as être), and characterized by [...] semantic density [...]. (Waugh and Monville-Burston, 1986: 874)[20]

Engel (1989) also found in her study of past tenses with *être* that a number of other factors were at play in the choice between *passé simple* and *passé*

---

[19]Bybee argues that this paradox is to do with the lexical diffusion of phonetic vs. morphophonemic change; see 2003: 621 for full explanation.

[20]See Waugh and Monville-Burston, 1986: 873-874 for discussion of 'semantic density' in relation to the simple past.

compose, i.e. *fut* and *a été*. While Engel's study is also principally concerned with written language (journalistic texts), she also carries out cloze tests with francophone participants, and the results can help shed light on what may be happening in the present study. In Engel's results, *être* was the only verb to be found with more tokens of the *passé simple* (PS) than the *passé composé* (PC) in the journalistic passages. In the cloze tests, Engel found that *fut* seemed to be used in specific, well-defined contexts, whereas *a été* could often be replaced by other tense forms. The factors that seemed to be involved in the preference for *être* in the PS included phonological (e.g. number of syllables, phonological quality), stylistic (word order, inversion) and syntactic factors (e.g. ce + *être*, passive constructions, PC with *être* as auxiliary).[21] The most relevant for our purposes are the first group, phonological factors. Relative number of syllables was a highly significant factor, in that shorter forms were strongly preferred – which in the case of *être* meant the PS form, *fut*. Engel notes that 'les effets de contraste [...] resortissent plus au style des auteurs qu'un emploi automatique chez la plupart des francophones instruits' (1989 : 6); i.e. it is not necessarily a deciding factor for the average speaker, though in her results these effects did play a role. In relation to our results here, for future temporal reference, the IF forms *sera* and *aura* are shorter than their PF equivalents *va être* and *va avoir*.[22]

Another factor investigated by Engel is *qualité phonologique*, phonological quality, which includes assonance, alliteration and avoidance of cacophony (by 'l'interruption d'une séquence de mots à voyelle intiale'; 1989: 7). This last may be relevant here: more 'euphonic' forms (i.e. those avoiding cacophony) were preferred in Engels' study, and for *être*, this was again the PS form *fut*. Here, the IF forms have an advantage over the PF forms, in that they phonologically more euphonic, and avoid hiatus, i.e. two vowels coming together in *va être* and *va avoir*.

Further investigation is needed to confirm number of syllables and phonological quality (cacophony/avoidance of vowels in hiatus) as factors in the choice of FTR forms with *avoir* and *être*, but Bilger's (2001) study of FTR in oral French contains some insights which lend weight to the hypothesis that these might be significant. Bilger found that IF was highly likely to appear with the verbs *être*, *avoir*, *pouvoir*, *devoir*, *falloir* and *faire* – i.e. the forms *sera*, *aura*, *pourra*, *devra*, *faudra*, *fera*. Bilger's explanation is more concerned with verb type and lexical semantics. She notes:

> D'une part, la valeur stative ou non-stative des verbes semble effectivement sélectionner l'emploi du futur périphrastique ou du futur simple ; d'autre part, au regard de leur fréquence, le sémantisme des verbes modaux semble être en parfaite adéquation avec la « tension modale » du futur simple et du conditionnel. (2001: 185)

---

[21]Engel notes that phonological factors are important even in a study of written language, since 'les écrivains professionels tiennent bien compte du 'son' de leurs phrases' (1989: 6).

[22]In Hansen's study of schwa/e-caduc, *sera* was virtually always pronounced without schwa, i.e. the single-syllable pronunciation [sʁa] would be the norm (Hansen, 1994: 38).

Table 8. 'Emplois stéréotypes' noted in Bilger (2001: 188)

Futur simple	
Type de sujet	**Lexique verbal**
On ça/ce	aura, sera, pourra, fera, devra
Futur périphrastique	
Type de sujet	**Lexique verbal**
il ya… qui va on a… qui va c'est…qui va	être, faire, voir, venir, avoir

However, it is surely of note that all the preferred forms here are shorter and more easily retrieved due to their high frequency. Bilger hints at this when she notes (in relation to the second part of her study, based on the GARS corpus): 'le futur périphrastique utilise une variété plus grande de lexique verbal que le futur simple qui concentre ses emplois sur une petite série de verbes, et notamment dans les deux cas sur *être* et *avoir* ainsi que sur les verbes modaux *pouvoir et falloir*' (2001: 187). Bilger (2001: 188) describes certain uses as 'emplois stéréotypes', noting the following strong tendencies set out in Table 8.

Additional support for the hypothesis that these verbs' IF forms are in some sense stereotypical or formulaic comes from B Lorenz's 1989 oral corpus study. Lorenz finds that while PF tends to combine with dynamic, lengthy and reflexive verbs, IF tends to be found with stative and short verbs. Thus there is a small group of highly frequent IF forms which are also shorter and more euphonic than their PF counterparts. They seem to be more easily accessed than PF forms for certain verbs, including *avoir* and *être*. These factors all contribute to higher rate of IF selection.

Two recent studies offer further evidence to support this hypothesis: Côté (2018) and Tremblay et al. (2019). Côté (2018) looks at FTR in L2 French. The authors hypothesize that, given the analytical and (they argue) cognitively simpler construction of the periphrastic future, this will be the preferred form among L2 speakers in their study, especially for irregular verbs, whose more complex morphology, it is predicted, will discourage the use of IF. This is indeed what is found overall. However, *avoir* and *être* did not conform to this pattern: participants used IF with these two verbs more frequently than with any other verb, regular or irregular (2018: 540). The authors attribute this to learners' earlier exposure to, and thus greater familiarity with, these two verbs, as well as to their high frequency overall. In their study of FTR in text messages, Tremblay et al. (2019) also found that verb type was a significant factor, with *avoir*, *être* and modal verbs being much more likely to appear with IF.[23] They

---

[23]Tremblay et al. (2019) include the [futurate] present in their analysis, and indeed this is found to be the most common form used to refer to future time in their corpus (see Tremblay et al. 2019 for in-depth discussion of the significance of this novel finding).

argue that this is related to morphological complexity: the more irregular the paradigm, the more likely the use of IF. The authors suggest that this effect may be due to accessibility of forms in the mental lexicon, with highly frequent irregular forms being more easily retrievable.

6. CONCLUSION

This investigation set out to investigate the distribution of FTR forms with two verbs in particular, *avoir* and *être*, because in some previous studies, these (and a handful of other verbs) have been identified as exhibiting different distributional patterns to the majority of verbs. The results showed that *avoir* and *être* strongly prefer IF, though this tendency has weakened somewhat for *être* in the more recent data. It is of course possible to take lexical differences into account when using mixed-effects statistical models to evaluate the effect of different factors on selection of variants. Nevertheless, to do this does not further our understanding of *why* this might be the case; it merely ensures that the statistical model is as accurate as possible regarding the effect of the independent factor groups investigated (see Johnson, 2009, for discussion of the use of mixed effects in variable rule analysis of linguistic data).

Complementing previous studies of FTR in the variationist paradigm, and others which take a different approach (e.g. Abouda and Skrovec, 2015, 2017), this study has shed some light on what lies behind the somewhat exceptional behaviour of *avoir* and *être*. In doing so, we have underlined the importance not only of taking lexical differences into account when conducting analyses which throw all verbs (or other word categories) in together, but also of using a combination of quantitative and qualitative analysis to explore this variation further.

REFERENCES

Abouda, L. and Skrovec, M. (2015). Du rapport entre formes synthétique et analytique du futur. Étude de la variable modale dans un corpus oral micro-diachronique. *Revue de Sémantique et Pragmatique*, **38**: 35–57.

Abouda, L. and Skrovec, M. (2017). Alternance futur simple/futur périphrastique : variation et changement en français oral hexagonal. *Revue de Sémantique et pragmatique*, **41–42**(41–42), pp. 155–179.

Antonini, A. (1753). *Principes de la grammaire françoise, pratique et raisonnée*. Paris: Duchesne.

Armstrong, N. (2001). *Social and Stylistic Variation in Spoken French: a Comparative Approach*. Amsterdam/Philadelphia: John Benjamins.

Ashby, W. (1981). The loss of the negative particle ne in French: a syntactic change in progress. *Language*, **57**: 674–687.

Beeching, K. (2002). *Gender, Politeness and Pragmatic Particles in French*. Amsterdam/Philadelphia: John Benjamins.

Bilger, M. (2001). Retour sur le « futur » dans les corpus du français parlé. *Recherches sur le français parlé*, **16**: 177–188.

Blanche-Benveniste, C. (1977). L'un chasse l'autre, le domaine des auxiliaries. *Recherches sur le français parlé*, **1**: 100–148.

Blanche-Benveniste, C., Bilger, M. and Rouget, C. (1990). *Le français parlé: Études grammaticales*. Paris: Centre National de Recherches Scientifiques.

Blas Arroyo, J. L. (2008). The variable expression of future tense in Peninsular Spanish: the present (and future) of inflectional forms in the Spanish spoken in a bilingual region. *Language Variation and Change*, **201**: 85–126.

Blondeau, H. (2006). La trajectoire de l'emploi du futur chez une cohorte de Montréalais francophones entre 1971 et 1995. *Canadian Journal of Applied Linguistics/Revue canadienne de linguistique appliquée*, 9.2: 73–98.

Blondeau, H. and Labeau, E. (2016). La référence temporelle au futur dans les bulletins météo en France et au Québec : regard variationniste sur l'oral préparé. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 61.3: 240–258.

Bres, J. and Labeau, E. (2013). About the illustrative use of the aller + Infinitive periphrasis in French. In: E. Labeau and J. Bres (eds), *Evolution in Romance Verbal Systems*. Bern: Peter Lang, pp. 171–202.

Bybee, J. (2003). Mechanisms of change in grammaticization: the role of frequency. In: B. D. Joseph and R. D. Janda (eds), *The Handbook of Historical Linguistics*. Oxford: Blackwell, pp. 602–623.

Bybee, J. and Thompson, S. (1997). Three frequency effects in syntax. *Proceedings of the Twenty-Third Annual Meeting of the Berkeley Linguistics Society*, **23**: 378–388.

Cheshire, J. (1987). Syntactic variation, the linguistic variable and sociolinguistic theory. *Linguistics*, **25**: 257–282.

Comeau, P. (2015). Vestiges from the grammaticalization path: the expression of future temporal reference in Acadian French. *Journal of French Language Studies*, 25.3: 339–365.

Comeau, P., King, R. and LeBlanc, C. L. (2016). The Future's path in three Acadian French varieties. *University of Pennsylvania Working Papers in Linguistics*, 22.2: 21–30.

Comeau, P. and Villeneuve, A.-J. (2016). Future temporal reference in French: An introduction. *Canadian Journal of Linguistics*, 61.3: 231–239.

Confais, J.-P. (1995). *Temps mode aspect: Les approches des morphèmes verbaux et leurs problèmes à l'exemple du français et de l'allemand*. Toulouse: Presses Universitaires du Mirail.

Côté, S. (2018). L'effet de la (ir)régularité verbale sur le choix de variante du futur en français L2. *Canadian Modern Language Review*, 74.4: 523–547.

Coveney, A. (1996). *Variability in Spoken French: A Sociolinguistic Study of Interrogation and Negation*. Exeter: Elm Bank.

Coveney, A. (2007). Grammatical variation in French. *Nottingham French Studies*, **46**: 100–118.

Damourette, J. and Pichon, E. (1911). *Des mots à la pensée : essai de grammaire de la langue française*. Paris: D'Artrey.

Edmonds, A., Gudmestad, A. and Donaldson, B. (2017). A concept-oriented analysis of future-time reference in native and near-native Hexagonal French. *Journal of French Language Studies*, 27.3: 381–404.

Emirkanian, L. and Sankoff, D. (1986). Le Futur simple et le futur périphrastique dans le français parlé. In: J.-C. Bouvier (ed.), *Morphosyntaxe des langues romanes. Actes du XVIIe Cong. Internat. de Ling. and Philol. Romanes, IV (Aix-en-Provence, 29 août-3 sept. 1983)*. Aix-en-Provence: Université de Provence, pp. 395–407.

Engel, D. M. (1989). Les temps passes de être. *Revue Romane*, **241**: 3–12.

Fleischman, S. (1982). *The Future in Thought and Language : Diachronic Evidence from Romance*. Cambridge: Cambridge University Press.

François, D. (1974). *Français parlé : Analyse des unités phoniques et significatives d'un corpus recueilli dans la région parisienne (2 vols)*. Paris: Société d'études linguistiques et anthropologiques de France.

Gadet, F. (1997). La variation en syntaxe. *Langue Française*, **11**: 126–132.

García, E. (1985). Shifting variation. *Lingua*, **67**: 189–224.

Gougenheim, G. et al. (1964) *L'élaboration du français fondamental (1er degré): Étude sur l'établissement d'un vocabulaire et d'une grammaire de base*. Paris: Didier.

Greenberg, J. (1978). *Universals of Human Language, Vol. IV: Word Structure*. Stanford: Stanford University Press.

Grimm, R. D. (2010). A real-time study of future temporal reference in spoken Ontarian French. *University of Pennsylvania Working Papers in Linguistics*, 16.2: 16–23.

Grimm, R. and Nadasdi, T. (2011). The Future of Ontario French. *Journal of French Language Studies*, 21.2: 173–189.

Hansen, A. B. (1994). Etude du E caduc – stabilisation en cours et variations lexicales. *Journal of French Language Studies*, 4.1: 25–54.

Imbs, P. (1960). *L'Emploi des temps verbaux en francais moderne*. Paris: Klincksieck.

Imbs, P. (1968). *L'emploi des temps verbaux en français moderne: Essai de grammaire descriptive*. Paris: Klincksieck.

Jeanjean, C. (1988). Le Futur simple et le futur périphrastique en français parlé: Étude distributionnelle. In: C. Blanche-Benveniste, A. Chervel and M. Gross (eds), *Grammaire et histoire de la grammaire*. Aix-en-Provence: Université de Provence, pp. 235–257.

Johnson, D. E. (2009). Getting off the GoldVarb standard: introducing Rbrul for mixed-effects variable rule analysis. *Linguistics and Language Compass*, 3.1: 359–383.

Kahn, F. (1954). *Le système des temps de l'indicatif chez un Parisien et chez une Bâloise*. Geneva: Droz.

King, R. and Nadasdi, T. (2003). Back to the future in Acadian French. *Journal of French Language Studies*, **13**: 323–337.

Labeau, E. (2015). Il était une fois le passé simple…. *Journal of French Language Studies*, 25.2: 165–187.

Lavandera, B. R. (1978). Where does the sociolinguistic variable stop? *Language in Society*, **7**: 171–182.

Lesage, R. and Gagnon, S. (1992). Futur simple et futur périphrastique dans la presse québécoise. In: A. Chrochetière, J.-C. Boulanger and C. Ouellon (eds), *Actes du XV Congrès International des Linguistes*. Quebec: Presses de L'Université Laval, pp. 367–370.

Lindqvist, C. (1979). *L'Emploi temporel dans la complétive au subjonctif introduite par un temps du passé en français contemporain*. Uppsala: Acta Universitas Upsaliensis.

Lorenz, B. (1989). *Die Konkurrenz zwischen dem futur simple und dem futur périphrastique im gesprochenen Französisch der Gegenwart*. Münster: Kleinheinrich.

Poplack, S. (2001).Variability, frequency, and productivity in the irrealis domain of French. In: J. Bybee and P. Hopper (eds). *Frequency and the Emergence of Linguistic Structure (Typological Studies in Language 45)*. Amsterdam/Philadelphia: John Benjamins, pp. 405–428.

Poplack, S. and Dion, N. (2009). Prescription vs. praxis: The evolution of future temporal reference in French. *Language*, 85.3: 557–587.

Poplack, S., Lealess, A. and Dion, N. (2013). The evolving grammar of the French subjunctive. *Probus: International Journal of Latin and Romance Linguistics*, 25.1: 139–195.

Poplack, S. and St-Amand, A. (2007). A real-time window on 19th-century vernacular French: the récits du français québécois d'autrefois. *Language in Society*, 36.5: 707–734.

Poplack, S. and Tagliamonte, S. (2001). *African American English in the Diaspora: Tense and Aspect*. Malden: Blackwell.

Poplack, S. and Turpin, D. (1999). Does the futur have a future in (Canadian) French? *Probus: International Journal of Latin and Romance Linguistics*, 11.1: 133–164.

Roberts, N. S. (2012). Future temporal reference in hexagonal French. *University of Pennsylvania Working Papers in Linguistics*, **182**: 97–106.

Roberts, N. S. (2016). The future of Martinique French: The role of random effects on the variable expression of futurity. *Canadian Journal of Linguistics*, **61**: 286–313.

Romaine, S. (1984). On the problem of syntactic variation and pragmatic meaning in sociolinguistic theory. *Folia Linguistica: Acta Societatis Linguisticae Europaeae*, 18.3–4: 409–437.

Sankoff, D. (1988). Sociolinguistics and syntactic variation. In: F. J. Newmeyer (ed.) *Linguistics: The Cambridge Survey Vol. 4*. Cambridge: Cambridge University Press, pp. 140–161.

Sankoff, G. (1980). Above and beyond phonology in variable rules. In: G. Sankoff (ed.), *The Social Life of Language*. Philadelphia: University of Pennsylvania Press, pp. 81–94.

Söll, L. (1983). De la concurrence du futur simple et du futur proche en français moderne. In: F.-J. Hausmann (ed.), *Etudes de grammaire française descriptive*. Heidelberg: Julius Groos, pp. 16–24.

Sundell, L.-G. (1991). *Le temps futur en français moderne*. Uppsala: Textgruppen i Uppsala AB.

Tagliamonte, S. (2002). Comparative sociolinguistics. In: Chambers, J. K., Trudgill, P., and Schilling-Estes, N. (eds), *Handbook of Language Variation and Change*. Malden and Oxford: Blackwell, pp. 729–763.

Tagliamonte, S. (2012). *Variationist Sociolinguistics : Change, Observation, Interpretation*. Oxford: Wiley-Blackwell.

Tremblay, M., Blondeau, H. and Labeau, E. (2019). Texting the future in Belgium and Québec : present matters. *Journal of French Language Studies*, 30.1: 73–98.

Tristram, A. (2014). *Variation and Change in French Morphosyntax: the Case of Collective Nouns*. Cambridge: Legenda.

Vet, C. (1980). *Temps, aspects et adverbes de temps en français contemporain: Essai de sémantique formelle*. Geneva: Droz.

Villeneuve, A.-J. and Comeau, P. (2016). Breaking down temporal distance in a continental French variety: Future temporal reference in Vimeu. *Canadian Journal of Linguistics*, 61.3: 314–336.

Wagner, S. E. and Sankoff, G. (2011). Age grading in the Montréal French inflected future. *Language Variation and Change*, 23.3: 275–313.

Wales, L. (2014). The case of the invisible postman: the current status of the French future tense. In: R. Pensalfini, M. Turpin and D. Guillemin (eds), *Language Description Informed by Theory*. Amsterdam: John Benjamins, pp. 319–335.

Wales, M. L. (1983). The semantic distribution of « aller » + infinitive and the future tense in spoken French. *General Linguistics*, 23.1: 19–28.

Waugh, L. R. and Monville-Burston, M. (1986). Aspect and discourse function: the French simple past in newspaper usage. *Language*, 62.4: 846–877.

Weinreich, U., Labov, W. and Herzog, M. I. (1968). *Empirical Foundations for a Theory of Language Change*. Texas: University of Texas Press.

Winford, D. (1996) The problem of syntactic variation. In: J. Arnold and E. Al (eds), *Sociolinguistic Variation: Data, Theory, Analysis*. Selected Papers from NWAVE 23. Stanford, CA: CSLI, pp. 177–192.

APPENDIX
Frequency calculations for *avoir* and *être*

The frequency of the infinitive form of each verb was calculated for the ESLO corpus as a rough guide (it is not possible in ESLO to construct a simple search query which will capture all inflected forms). As a comparison, frequency in the FRANTEXT corpus of written French was also calculated, both for the infinitive alone, and for the infinitive plus inflected forms (the FRANTEXT interface allows for simpler searching for inflected forms).[24]

Table 9. Frequency calculations for avoir and être

		Absolute freq.	Total no. words	Freq. in corpus	Freq. per 1000 words
être	ESLO (infinitive)	7,435	2,447,632	0.003038	3.04
	FRANTEXT (infinitive)	3,770	1,767,529	0.002133	2.13
	FRANTEXT (inc inflected forms)	41,151	1,767,529	0.023282	**23.28**
avoir	ESLO (infinitive)	3,623	2,447,632	0.001480	1.48
	FRANTEXT (infinitive)	1,424	1,767,529	0.000806	0.81
	FRANTEXT (inc inflected forms)	31,300	1,767,529	0.017708	**17.71**

[24] <https://www.frantext.fr/> [last accessed February 2019].