

**THE SOCIOLINGUISTICS OF LEXICAL VARIATION  
IN STANDARD FRENCH:  
A DIACHRONIC PERSPECTIVE**

**KATE BEECHING**

**Abstract**

*In his volume on Parisian French, Lodge (2004) highlights the importance of lexical variation in French, finding it anomalous that sociolinguists, in the Labovian tradition at least, have neglected this highly salient aspect of speaker identity. As he says (2004: 228), by comparison with phonetics and grammar, the lexicon has been largely neglected outside the extensive inventories of dialect-specific words established in, for example, Rézeau (2001); some work has been conducted on Canadian French (e.g. Nadasdi, Mougeon and Rehner, 2008, on automobile) and on hexagonal French by Armstrong (e.g. 2001, Chapter 6, on (pre)adolescents' use of non-standard lexical items) but a systematic study of lexical variation in contemporary spoken French is still lacking.*

*This paper attempts to go some way to closing this gap by investigating the use of fifteen lexical doublets such as *boulot/travail* in three corpora of spoken French spanning the period from 1968 to 2002.*

*By creating a metavariable combining rates of occurrence of the fifteen terms, it was possible to demonstrate that the distributional frequency of colloquial items has risen in both real and apparent time in the last 40 years. Some caveats are issued with respect to the comparability of the corpora and the influence of the semantics of the terms selected (rates of *boulot*, *rigolo* and *copain* are higher for young people, while older interviewees tend to refer to *bouquin* and *bouquinier* to a greater extent). Overall, it appears that rates of occurrence of colloquial lexical items are rising, suggesting a move towards greater positive politeness and a reduction in social distance, echoed, too, in an increase in *tutoiement*. Persistent economic divisions, however, and the fact that the least educated are less prone to style-shift than the more educated, reveal the fictive nature of what might appear to be a move towards greater social equality.*

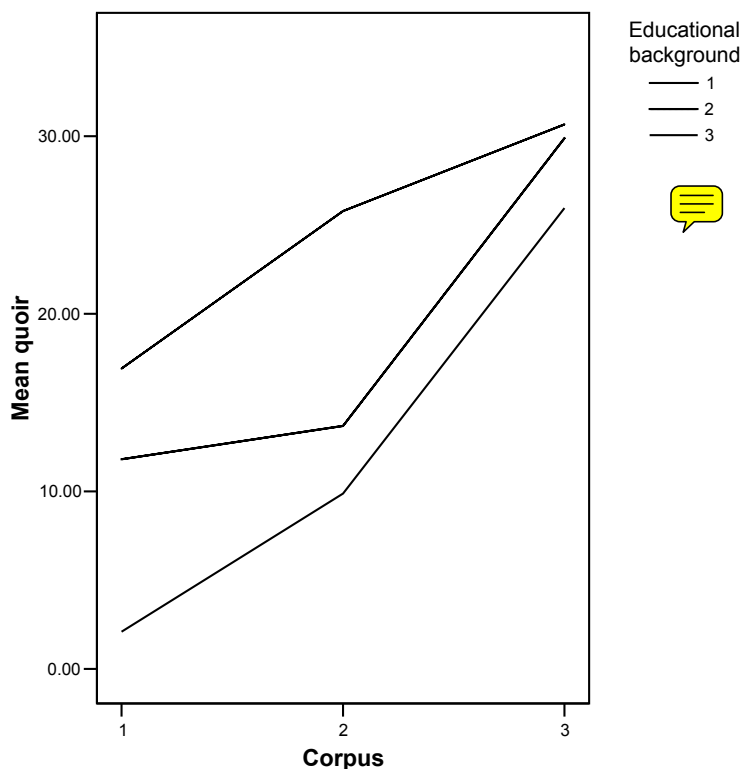
## 1. Introduction

In a series of papers in the late 1980s to the late 1990s (Lodge, 1989; 1997; 1999), Lodge problematises both the labels given by lexicographers to non-standard or colloquial terms, such as *fam.* (familiar), *pop.* (popular), *vulg.* (vulgar) and *arg.* (slang) and the prescriptive approach to the use of these terms, which tended to categorise them according to their ‘fitness for use in polite society’. Guiraud (1956; 1973; 1978; 1985; 1992) is cited as providing an analysis of the sociolinguistic function of slang – but his analyses largely collapse into a series of social stereotypes. *L’argot* and *le français populaire* are linked, according to Guiraud, to the communicative needs of *le peuple*, who favour an emotive and exaggerated manner of expressing themselves, referred to by Guiraud (1992: 40) as a highly negative characteristic: ‘La conversation des gens sans culture baigne dans l’affectivité’. Lodge challenges the stigma attached by Guiraud to the use of such terms and links their usage not to the social category *le peuple* but rather to social situations and to positive politeness.

This echoes the tenor of my own recent work on pragmatic particles in French, which appear to be on the increase and to be used by speakers from all social strata for reasons of politeness also explored by Wheeler (1994). Figure 1, reproduced from Beeching (2007a: 145) shows the extent to which usage of post-rhematic *quoi* (exemplified in *C’est superbe, quoi*) has increased and spread from the less to the more educated members of the French population in the early twenty-first century. In the 1968 corpus, rates of *quoi* usage are highest in the least educated group with very low rates of usage in the most educated members of the population. Rates of *quoi* usage have soared in the most educated group since 1968 to the point where, in the 2002 corpus, there is virtual convergence amongst all social groups in the use of this marker.

Wheeler (1994: 160) addresses the question raised by Milroy (1992: ix-x) as to why changes in the prestige norm seem to originate in ‘lower-status’ varieties, rather than in elite ones. Wheeler argues that speakers adopt a casual style in order to implement Positive Politeness: ‘To do otherwise would be to invite the hearer to infer that the speaker evaluated the relationship as less than satisfactorily solidary’. As there is a payoff in terms of social approval in being slightly more informal (than one’s hearer, than one’s parents, than the norm), a positive feedback loop is created whereby innovatory devices are created to mark intimacy. Wheeler goes on to suggest (1994: 145) that working class cultures are differentiated from middle-class cultures (in the U.K., at least), tending to favour positive and negative politeness, respectively. Beeching (2007a; 2007b; 2009) provides substantiation, for French, for the runaway positive feedback loop with respect to Positive Politeness and casual speech described by

Wheeler (1994). As we have seen, rates of stigmatised post-posed *quoi* rose dramatically between 1968 and 2002 and working-class and middle-class rates converged during this period, suggesting a relative democratisation of French society. What is more, hedging and hesitatory usages of discourse marking *bon* also appear to have increased in distributional frequency. This leads me to hypothesise that a similar situation may pertain with respect to the lexicon.



**Figure 1. Average rates of post-rhematic *quoi* 1968-2002**

**quoir** average rates of occurrence of post-rhematic *quoi* per 10,000 words

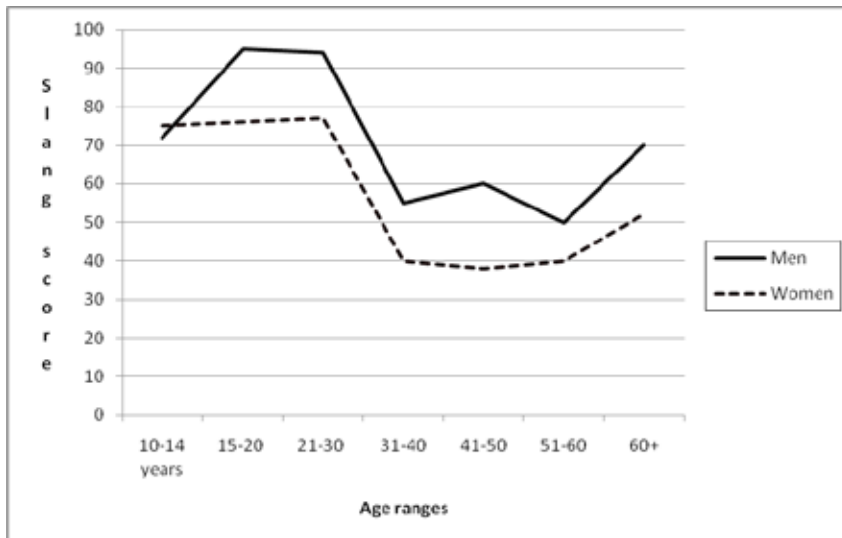
**Corpus:** **1** – 1968 *Enquête Sociolinguistique d’Orléans*; **2** – 1988 Beeching Corpus; **3** – 2002 *Corpus de Référence du Français Parlé*

**Educational Background:** **1** – left school by age 16; **2** – left full-time education at 18; **3** – Graduated from university with a degree or degrees

In his 1999 paper, Lodge reports the results of a survey undertaken in Clermont-Ferrand ten years previously on the use of colloquial lexical items such as *bouquin*, *boulot*, *flic*, *bagnole*, *bouffer* and *con* along with other forms generally labelled in dictionaries as ‘popular’ or ‘vulgar’ such as *nana*,

*clope, gonzesse* etc. Interviewees were asked to say how frequently they used particular words – never, rarely, often or very often. In Question 1, they were asked this question with respect to intimates and in Question 2 with respect to people they did not know.

Lodge correlated the interviewees ‘slang score’ with sociolinguistic variables: the age, sex and social class of the speakers. The results for Question 1 (use with intimates) are reproduced in Figure 2.



**Figure 2. Correlation of ‘slang’-use with age and sex Lodge (1999: 360)**

As Lodge remarks, the oft-noted age and gender pattern, whereby men have greater recourse to colloquial terms than women and usage of such terms correlates with degree of involvement in the economic/professional market, is closely followed in the Clermont survey.

If Question 1 elicited informal or intimate style, Question 2 was assumed to elicit interviewees’ more careful style – scores here are charted in Table 1:

	Q1	Q2
I Cadres Supérieurs/Professions libérales	72.1	57.2
II Cadres moyens	65.3	43.7
III Employés	88.0	85.1
IV Ouvriers/Personnel du service	74.3	69.8

Lodge (1999: 360) notes two interesting correlations:

- (a) whereas the highest and lowest social groups give themselves roughly the same slang score, it is the second highest group that claims the lowest slang score, reflecting the well-documented tendency of the lower middle class to linguistic insecurity and hypercorrection; (b) if there is a difference between the behaviour of the upper and lower social groups, it lies in the tendency of the upper groups to shift style more sharply between informal situations (responses to Question 1) and formal situations (responses to Question 2) than the lower groups.

This latter observation is reflected in Buson's (2009) study of French preadolescents' perceptions of stylistic variation. Buson set up two indices R<sub>n</sub> (normative criteria) and R<sub>i</sub> (interactional criteria) in her analysis of preadolescents' comments on a set of stylistically varied answer phone messages – she observes (2009: 156) that:

the ability to refer to interactional characteristics in interpreting stylistic varieties (such as familiarity with the interlocutor or the communicative intentions of the speaker) is correlated with social status ... upper-class preadolescents have higher R<sub>i</sub> scores than lower-class preadolescents.

That is to say that the lower-class group related more familiar forms with lower-class speakers and more 'formal' variants with higher-class groups, whereas the upper-class preadolescents were less likely to correlate style with class but rather, more flexibly, with the situation.

This is echoed in part in Lodge's (1999: 365) conclusion that:

...since all speakers, regardless of their place in the social hierarchy, have some recourse to this sort of vocabulary on certain occasions, as an expression of positive politeness, it is preferable to use speech situation as the more basic determinant of its use. The fact that certain types of speaker find themselves in positive politeness situations more often than others does not detract from this. Speech situation is a more basic and universal determinant of the nature of colloquial vocabulary than the social origins of its supposed speakers.

## **2. The present study**

Lodge (1999) notes that the Clermont-Ferrand survey did not investigate the informants' actual use of non-standard vocabulary as it bore only on their perceived use of such items. He highlights previous studies in which certain groups over-report their use of vernacular items (typically young males), while others tend to do the reverse (typically middle-aged females). He remarks (1999: 359):

An investigation of actual use would have provided more valuable information. However, access to statistically significant numbers of lexical variables of informal language use is not easy.

With the development of sociolinguistically stratified corpora of spoken language from the *Enquête Sociolinguistique d'Orléans* in 1968 to the *Corpus de Référence du Français Parlé* up till 2002, we are in a slightly better position to chart not only age, sex and class differences in the usage of non-standard vocabulary but also to trace diachronic developments in this area. The present study aims to do just that.

The difficulties associated with studying lexical variation are well-known:

- a. single vocabulary items are infrequent by comparison with phonological items;
- b. lexical variants, unlike phonological and some morphological ones, are frequently said not to be semantically equivalent and thus are not 'variant forms' of a given variable;
- c. lexical differences between varieties are easily recognisable and are highly salient to ordinary (non-linguistically trained) speakers;
- d. there are fashions in lexical usages and items can be short-lived.

In response to a) and with the aim of obtaining statistically robust results, a decision was taken to investigate a small sub-set of only fifteen lexical pairs (*bagnole*\*/*voiture*\*; *balade*\*/*promen*\*; *boss*\*/*travail*\*; *bouff*\*/*mang*\*; *boulot*/*travail*; *bouquin*\*/*livre*\*; *copain*\*/*ami*\*; *flic*\*/(*gendarme*\*/*policier*\*)<sup>1</sup>; *fric*/*argent*; *fringue*\*/*vêtement*\*; *\*gueule*\*/*bouche*; *mec*\*/*homme*\*; *\*merde*\*/*lennu*\*; *pif*/*nez*; *rigol*\*/*amus*\*). These lexical items are relatively frequent, most could be said to be semantically synonymous, and most are well entrenched in everyday speech and can be said to be stable, thus addressing, at least to some extent, points b) and d). Point c) is less easily addressed, but none of these forms is recognisably regional and the variation between the forms, stylistically, situationally or from the point of view of social stratification is precisely the variability that is under study.

A further response to the frequency problem posed by a) is to treat the set of variable lexical items as a group variable or 'metavariable', a solution

1 The asterisks indicate that the lemmatised version of the form was used, so that both singular and plural, verbal as well as nominal or adjectival forms, and derived forms, where appropriate, were retrieved from the corpora. The asterisk was used as a 'wild card' in the searches.

proposed by Armstrong (2001: 217). Details of the corpora consulted are listed at the end of the chapter.

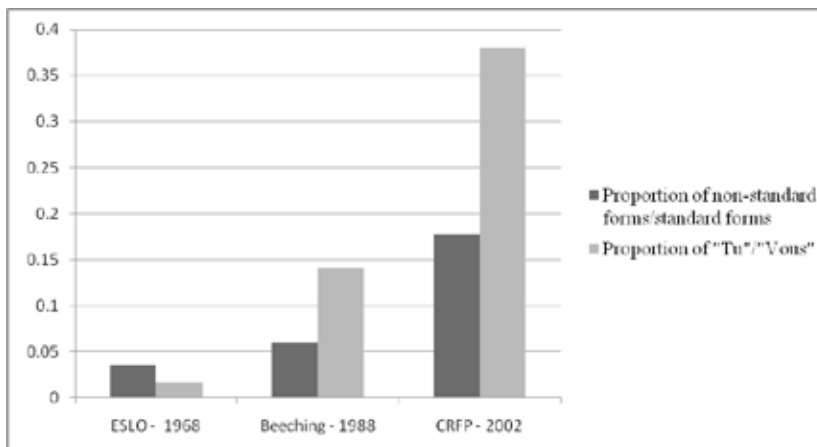
### 3. Results

#### 3.1 Diachronic factors: real-time data

Table 2 charts the raw number of occurrences of the ‘informal’ and ‘formal’ variants. Numbers of occurrences in these sociolinguistic interviews are fairly low, with *bossier*, *boulot*, *bouquin*, *copain* and *rigoler* proving the most frequent. *Gueule* appears to be more frequent than *bouche*, unsurprisingly given its ability to appear in derived forms which do not specifically relate to the semantic concept ‘mouth’, a point which will be taken up again below. By way of comparison, raw numbers of occurrences of *tu* and *vous* are included at the end of the table. These figures are raw figures and do not take account of the fact that the word count for each corpus varies. It is, however, immediately evident that the proportion of usage of *tu* with respect to *vous* has risen in these corpora when we compare the data from 1968 with 1988 and 2002.

<b>Table 2. Raw number of occurrences of informal/ formal variants in the three corpora investigated</b>			
	<b>ESLO (1968) 303, 357 words</b>	<b>Beeching (1988) 154, 357 words</b>	<b>CRFP (2002) 287, 482 words</b>
<i>bagnole</i> */ <i>voiture</i> *	0/69	1/37	0/51
<i>balade</i> */ <i>promen</i> *	1/27	1/23	8/12
<i>boss</i> */ <i>travail</i> *	0/395	1/286	20/269
<i>bouff</i> */ <i>mang</i> *	0/49	2/91	5/70
<i>boulot</i> / <i>travail</i>	3/420	15/171	43/232
<i>bouquin</i> */ <i>livre</i> *	8/119	2/35	20/65
<i>copain</i> */ <i>ami</i> *	22/158	17/70	47/105
<i>flic</i> */( <i>gendarme</i> * / <i>policier</i> *)	0/2	1/10	5/0
<i>fric</i> / <i>argent</i>	0/44	0/58	5/67
<i>fringue</i> */ <i>vêtement</i> *	0/12	0/7	2/19
* <i>gueule</i> */ <i>bouche</i>	4/5	7/2	13/3
<i>mec</i> */ <i>homme</i> *	0/58	3/108	13/45
* <i>merde</i> */ <i>ennu</i> *	5/38	1/12	13/12
<i>pif</i> / <i>nez</i>	1/2	0/2	1/10
<i>rigol</i> */ <i>amus</i> *	9/29	8/19	22/21
<b>TOTAL</b>	53/1427 (1480)	59/931 (990)	212/981 (1193)
<i>tu</i> / <i>vous</i>	103/6216 (6319)	318/1927 (2245)	981/1601 (2582)

Figure 3 displays the fact that the proportion of both the familiar lexical forms over more standard forms (computed as a metavariable) and that of *tu* over *vous* increases regularly in real time from 1968 to 2002, in so far as this is reflected in the three sociolinguistic corpora investigated.



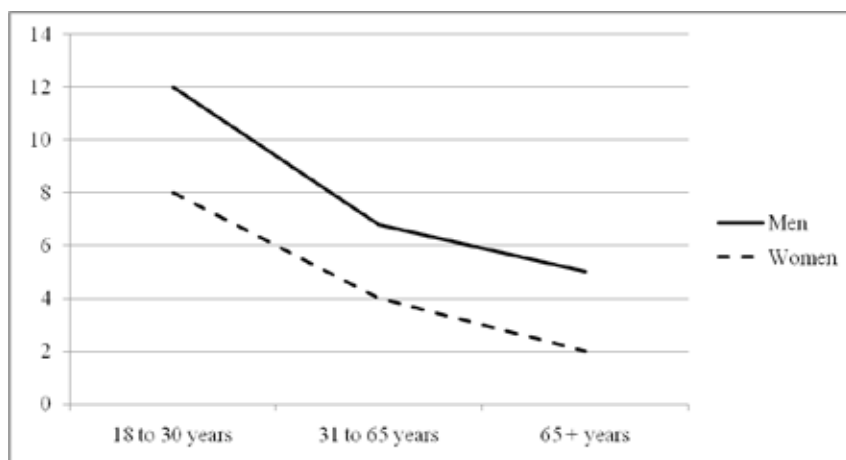
**Figure 3. Proportion of familiar/standard forms and of *tu/vous* in real time**

### 3.2 Apparent-time data: age and sex

Though the ESLO, Beeching and CRFP corpora are very similar in that all adopted a sociolinguistic interview approach, it might be argued that the recorded interview situation is perceived as less formal in style in the 21st century than in 1968 and that this could influence the number of colloquial items elicited. Another means of confirming the finding that rates of usage of 'informal' variants has risen is to investigate the phenomenon in apparent time, using the three generations of speakers in the (2002) *Corpus de Référence*, the 18-30 year-olds, the 31-65 year-olds and the over 65s. Average rates of occurrence of the familiar forms per 10,000 words are displayed in Figure 4, which subdivides the data according to sex. The figure appears to confirm Lodge's (1999) Clermont perceptual data that men use more familiar forms than women, though the difference in rates of usage is not statistically significant. Young people use significantly more familiar forms than the middle-aged and older generations, though the difference between the middle-aged and older generation is not statistically significant. It could be argued that the difference between the younger generation and the others is a case of 'age-grading' (a feature of the young – and not a feature of language change). However, in tandem with the real-



time data, presented in Figure 3, this is a further indication that the rate of usage of informal forms is increasing.

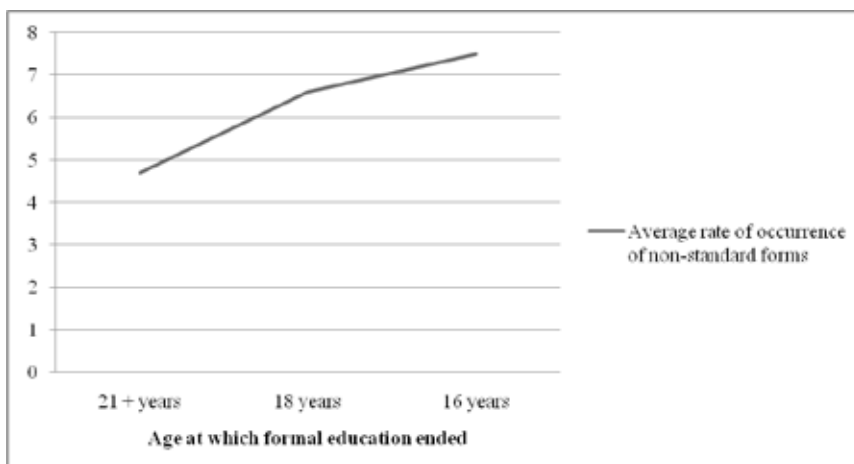


**Figure 4. Average rate of occurrence of familiar forms in the CRFP by age and gender (Statistical significance 2)**

### 3.3 Educational background

With respect to educational background, the comparison of rates of occurrence of familiar forms across speakers in the CRFP who completed their formal education with a university degree or degrees, at 18 or at 16 or below, is displayed in Figure 5. Rates of the familiar form metavariable rise regularly across the social classes, those with least schooling having a rate of 7.25 occurrences per 10,000 words while those with most score 4.6 per 10,000 words. These differences do not reach statistical significance and provide empirical support to Lodge's affirmation cited above that French colloquial vocabulary can be found in the communicative requirements of particular speech situations, which apply in varying degrees to all speakers of the language. 'That is not to say [however] that there is no correlation at all between social stratification and the use of colloquial vocabulary' (1999: 365). It appears that in the private interviews which feature in the *Corpus de Référence*, there is tentative evidence that less educated speakers use more colloquial forms than more educated speakers.

2 Level of statistical significance: Men/Women – no significance ( $z=-.266$ ; Asymp. sig.= .790) Age band 1/3 significant ( $z=-2.672$ ; Asymp.sig.=.008); 1/2 significant ( $z=-2.335$ ; Asymp.sig.= .02); 2/3 not significant ( $z=-.771$ ; Asymp.sig.= .441)



**Figure 5. Average rate of familiar forms in apparent time in the *Corpus de Référence du Français Parlé* according to educational background (Statistical significance?)**

#### 4. The metavariable and the role of semantics

As explained above, the metavariable is made up of all fifteen lexical pairs and it is of some interest to unpackage the usage made of some of the separate lexical items by the different age-groups and with different educational backgrounds. Figure 6 charts rates of average usage of eight of the more frequently occurring individual forms by members of the different age-groups.

<sup>3</sup> Level of statistical significance: no significance – for example: 16 year-old school-leavers/graduates (  $z=-.161$ ;  $Asymp.sig.=.872$ )

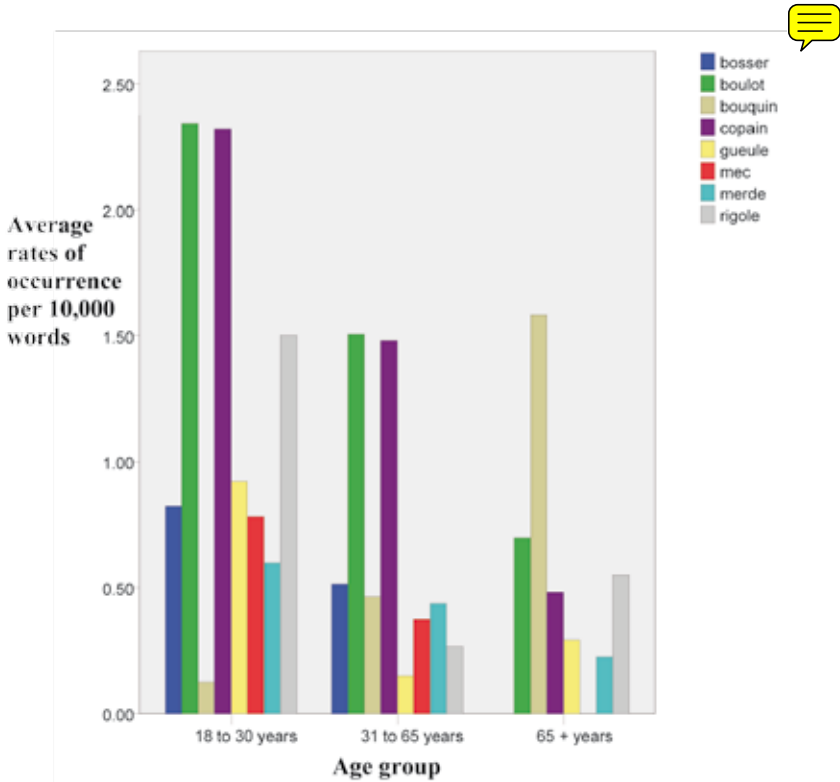
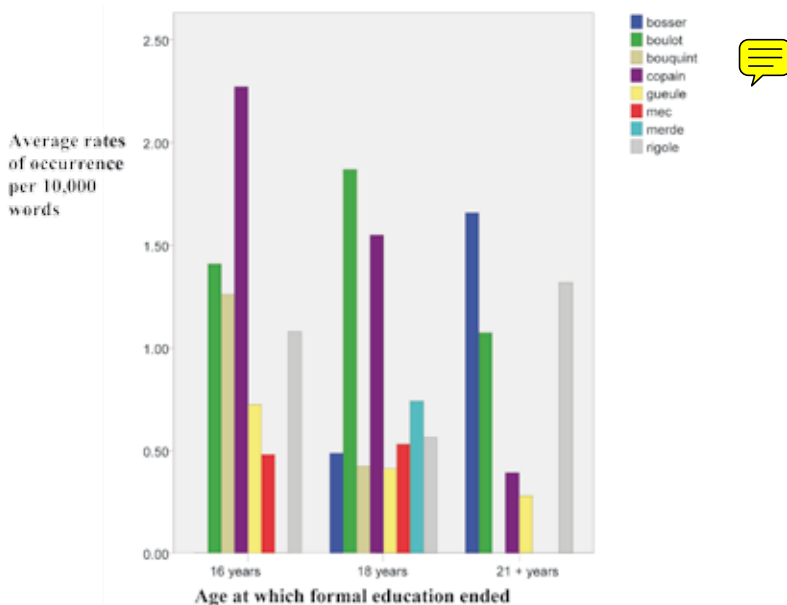


Figure 6. Individual forms and age group

The young speakers have higher rates of usage overall; the high-scoring ‘familiar’ forms amongst the young are *boulot* (work) and *copain* (mate). These might be said to be topics of specific interest for young people. Young people *rigolent* (have fun) and find things *rigolo* a great deal more than the older age-groups (though the elderly appear to have rather more fun than the middle-aged). The 65+ age-group refer to reading and books with higher rates of *bouquin(s)* and *bouquiner*. Young people, by contrast, make very sparse reference to books and reading. These data suggest that the semantic nature of the lexical item in itself (rather than its relatively formal or informal nature) might account for differences in rates of usage.

Rates of *\*merde\** usage, however, (including derived forms such as *emmerdeur* and *démerder*) might be expected to be similar across age-groups from a psychological or interest-oriented point of view (taking the stereotypically more pessimistic viewpoint of the older speakers into account) – and this colloquial term shows a higher rate of occurrence in the

young than in the old. Figure 7 shows rates of average usage of individual forms by speakers of differing educational backgrounds.



**Figure 7. Individual forms and educational background of the speakers**

Because rates of usage for individual items are low, it is difficult to extrapolate from the data and come to generalisable conclusions. However, it is striking that, while references are made to *boulot* in all three groups, *boss\** is entirely absent from the least educated group. The term *copain* is used a great deal more by the least educated than the most educated and so is *bouquin\**. *Mec* and *\*merde\** are not used by the most highly educated whereas the latter have the highest rates of *rigol\**.

All in all, results based on the use of the metavariable must be interpreted with caution, as it may be that the semantics of the lexical items selected to be included in the metavariable play an important role and could bias results. This seems, moreover, to be more salient with respect to age than to educational background. Differences between the age groups are statistically significant.

### 5. Further qualitative remarks

Forms such as *rigoler* have derived forms or idiomatic usages which were picked up by the lemma used to collect rates of occurrence. These

include verbal inflections and adjectival and nominal forms such as *rigolo* and *rigolade*, derived forms such as *gueuler*, *dégueuler*, *engueuler*, *se démerder* and *emmerder* and idiomatic usages such as *faire la gueule*, *gueule de bois*, *un petit merdeux* and *la groupie de merde*. Each of these expressions could be classified as colloquial to a greater and lesser degree. For this reason, conflating all such forms into an undifferentiated metavariable might also be considered to bias results.

The extent to which such forms are considered to be stigmatised is illustrated in one speaker's remark, seen in the following example:

+ parce qu'on peut avoir envie de faire la la figure j'allais dire  
de **faire la gueule** en parlant mal {rire} mais euh bon ben  
CRFP PRI-MO-2

The interviewee suggests that speakers may want to cut a dash by using 'bad' language and, in searching for the expression *faire (la) figure*, almost said *faire la gueule* by mistake. He laughs at the irony of almost using *faire la gueule* when referring to not 'talking proper', *faire la gueule* thus being considered to be over-colloquial. This example illustrates the fact highlighted above under **Methods** that stylistic differences are highly salient to ordinary (non-linguistically trained) speakers.

## 6. Historical and literary data

Our spoken data, drawn from corpora dating from 1968 to 2002, appear to show that rates of colloquial words have increased. We cannot, however, state this with a strong sense of certainty without considering data from earlier periods. As Labov has famously remarked (1994:11), historical linguistics suffers from 'making the best of bad data'. We do not have audio recordings of speakers before the 20th century and the best we can do is to investigate written texts and these typically do not contain highly colloquial spoken terms. The corpus collection Frantext has, however, the advantage of being very large with 210 million words. By way of a broad-brush indicative overview, rates of occurrence of *gueule* and *merde* are included in Table 3.

**Table 3. Distribution frequencies of *gueule* and *merde* in Frantext**

Scale: one asterisk represents a relative frequency of 50 millionths.

Gueule

	Absolute frequency	Relative frequency.	
1500-1599	11,142	1,953	*****
1600-1699	23,697	1,085	*****
1700-1799	33,292	940	*****
1800-1899	102,188	1,464	*****
1900-1999	150,571	1,532	*****
2000-2099	2,307	1,833	*****

Merde

	Absolute frequency	Relative frequency.	
1500-1599	38	6	***
1600-1699	26	1	*
1700-1799	9	0	
1800-1899	117	1	*
1900-1999	2,627	26	*****
2000-2099	65	51	*****

The *gueule* lemma, even in a literary corpus, appears to have maintained a strong presence across the centuries. In the early period, *gueule* is generally used to refer to the mouth of an animal and this accounts for its frequency of occurrence in sixteenth century texts. It has, however, maintained a strong position and merits more detailed study. The *merde* lemma on the contrary appears with any frequency in literary texts only in the 20th century and the rate of occurrence soars in the small number of texts already available for the 21st. This suggests that the use of this stigmatised term is considered more acceptable in recent years, even in the more formal context of the written language.

## 7. Conclusions

The study indicates that the distributional frequency of so-called colloquial items is rising in contemporary French. Of the reasons which we might propose for such a development, we can invoke social changes involving a general relaxation of formality both in France following May 1968 and world-wide, as Pinker (2007: 316) remarks. Social class hierarchies have become less rigid, and this is reflected in clothing, such as the ubiquitous wearing of jeans and tee-shirts, as well as in less formal modes of speech. Allied to this, we have seen an evolution in norms of politeness in France from a mode conforming to distance (use of *vous* in asymmetrical encounters) to one of greater camaraderie (greater use of *tu*).

Wheeler's (1994) insightful observation that more familiar forms allow for a stronger sense of connection or solidarity and that this creates a self-perpetuating feedback loop is borne out by the data we present here with respect to colloquial lexical forms.

It is worth pointing out, as do Smith (1996) and Armstrong and Smith (2002), that the blurring of hierarchies in France during the 1960s and 1970s and beyond has not been accompanied by a redistribution of wealth. As Armstrong and Smith remark (2002: 40), the change is symbolic and economic divisions remain:

Although no substantive change in the French economic structure has taken place during this period, most notably in terms of the distribution of wealth and income, important symbolic social changes have come about. After Italy, France continues to have the highest ratio of inequality (15:1) between the highest and lowest 10% of wage-earners in the OECD group of countries; at the same time French decision-makers now feel the need to adopt a consensual rather than a directive approach, and to emphasise solidarity rather than hierarchy. Social divisions, between the middle and working classes, men and women, young and old, have become blurred during this period, even though economic divisions are as sharp as before, or even sharper, as Smith points out (1996: 133-4)

In concluding their study of the decline of *ne* in contemporary French, Armstrong and Smith note the role of the greatly enhanced status of young people in the general informalisation of French over the last 40 years and the increasing acceptability of taboo terms.

With specific respect to colloquial vocabulary items, there are, however, in addition, intralinguistic causes for the rise in frequency which we have seen in the plethora of new forms derived from the original core items, such as *rigolo* and *rigolade*, *emmerder* and *merdeux*, *engueuler* and *dégueuler*. The histories, shifts of meaning, and propagation of such forms merit further study. Such derived forms permit a further presence and expansion of these items in common usage.

Finally, we must return to Lodge's (1999) stance with respect to the situationally or socially based nature of colloquial vocabulary and to explore this in relation to Buson's recent finding that children from less privileged backgrounds relate stylistic variants to speakers from particular social backgrounds while children from more privileged backgrounds relate them more easily to situational appropriacy. Lodge argued that, since all speakers can use colloquial terms in a context-sensitive way, regardless of their social background, speech situation is a more basic determinant of their use than the social origins of speakers. Such a stance chimes with post-modern constructionist views and Coupland's (2007) notion of 'styling'; rather than viewing particular forms as being associated with speakers

from particular social strata, linguistic variables are put to work creatively and dynamically. What we cannot help noting, however, is that Lodge's (1999) self-report findings, Buson's (2009) analysis of pre-adolescents' perceptions, and the outcome of the present study indicate that employing particular linguistic variables in creative and dynamic ways is less marked in less educated speakers. The ability to style-shift according to situation appears to be available to a far greater extent to the better educated or socially advantaged, a fact noted in Lodge (1993: 249 and 1999: 360). More highly educated speakers appear to be appropriating some of the colloquial vocabulary stereotypically associated with less educated speakers to hint at a social equality which may have, sadly, little basis in economic fact.

## REFERENCES

- ARMSTRONG, N. (2001) *Social and Stylistic Variation in Spoken French*. Amsterdam and Philadelphia: John Benjamins.
- ARMSTRONG, N. and SMITH, A. (2002) The influence of linguistic and social factors on the recent decline of French *ne*. *Journal of French Language Studies*, 12.1: 23-41.
- BEECHING, K. (2007a) Social identity, salience, and language change: The case of post-rhematic *quoi*. In: W. Ayres-Bennett and M.C. Jones (eds.) *The French language and questions of identity*. Oxford: Legenda, 140-149.
- BEECHING, K. (2007b) La co-variation des marqueurs discursifs *bon, c'est-à-dire, enfin, hein, quand même, quoi* post-rhématique et *si vous voulez*: une question d'identité? *Langue Française*, 154.2: 78-93.
- BEECHING, K. (2009) Sociolinguistic factors and the pragmaticalisation of *bon* in contemporary spoken French. In: K. Beeching, N. Armstrong and F. Gadet (eds.) *Sociolinguistic Variation in Contemporary French*. Amsterdam/Philadelphia: Amsterdam and Philadelphia: John Benjamins, 215-230.
- BUSON, L. (2009) French preadolescents' perceptions of stylistic variation: A contrastive sociolinguistic study. In: K. Beeching, N. Armstrong and F. Gadet (eds.) *Sociolinguistic Variation in Contemporary French*. Amsterdam and Philadelphia: John Benjamins, 141-158.
- COUPLAND, N. (2007) *Style, language and variation*. Cambridge: Cambridge University Press.
- GUIRAUD, P. (1992) [1956] [1973] [1985] *L'argot*. Paris: Presses universitaires de France (Que Sais-je?)



- GUIRAUD, P. (1978) [1965] *Le français populaire*. Paris: Presses universitaires de France (Que Sais-je?)
- LABOV, W. (1994) *Principles of linguistic change. Internal factors*. Oxford: Blackwell.
- LODGE, R. A. (1989) Speakers' perceptions of non-standard vocabulary in French. *Zeitschrift für Romanische Philologie*, 105.6: 427-44.
- LODGE, R. A. (1993) *French. From dialect to standard*. London: Routledge.
- LODGE, R. A. (1997) The pragmatics of slang. *Web Journal of Modern Language and Linguistics* Issue 2.
- LODGE, R. A. (1999) Colloquial vocabulary and politeness in French. *Modern Language Review* 94: 355-365.
- LODGE, R. A. (2004) *A sociolinguistic history of Parisian French*. Cambridge: Cambridge University Press.
- MILROY, J. (1992) *Linguistic variation and change*. Oxford: Blackwell.
- NADASDI, T., Mougeon, R. and Rehner, K. (2008) Factors driving lexical variation in L2 French: A variationist study of *automobile, auto, voiture, char* and *machine*. *Journal of French Language Studies*, 18: 365-381.
- PINKER, S. (2007) *The stuff of thought*. Harmondsworth: Penguin.
- RÉZEAU, P. (2001) *Dictionnaire des régionalismes de France: géographie et histoire d'un patrimoine linguistique*. Brussels: De Boeck/Ducoulot.
- SMITH, A. (1996) *A diachronic study of French variable liaison*. M Litt dissertation: University of Newcastle-upon-Tyne.
- VÉRONIS, J. (2005) Présentation du *Corpus de référence du Français Parlé*. *Recherches sur le français parlé* 18: 11-42.
- WHEELER, M. (1994) Politeness, sociolinguistic theory and language change. *Folia Linguistica Historica*, 15: 149-174.

### The corpora consulted

- FRANTEXT literary corpus: 210,000,000 words; 3,737 texts from the 16th.-20th. century.
- Enquête Sociolinguistique d'Orléans (1966-1970) : <http://bacharts.kuleuven.ac.be/elicop> . 109 hours (902,755 words transcribed); Sample of 24 speakers used.
- Beeching Corpus (1988-1990) : <http://www.uwe.ac.uk/hlss/llas/iclru/corpus.pdf>. 175 hours (155,000 words transcribed); 95 speakers.
- Corpus de Référence du Français Parlé "CRFP" (2002): <http://www.up.univ-mrs.fr/delic/crfp/> 40 towns in France, 400,000 words. See Véronis (2005); 82 speakers.

<b>Table 4. Constitution of the CRFP: Number of interviewees according to sex, education and age-range</b>						
			Age range			Total
			18 to 30	31 to 65	65 +	
Men	*Educ	up to 16 years	2	5	7	14
		up to 18 years	9	14	5	28
		up to 21 + years	1	0	2	3
Total			12	19	14	45
Women	*Educ	up to 16 years	4	0	3	7
		up to 18 years	12	9	1	22
		up to 21 + years	3	5	0	8
Total			19	14	4	37