

## Googling Toubon. Testing the effects of institutional French language purism.

Preprint of: Berthele, R. (2015). Googling Toubon. Testing the effects of institutional French language purism. In J. Daems, E. Zenner, K. Heylen, D. Speelman & H. Cuyckens (Eds.), *Change of Paradigms – New Paradoxes. Recontextualizing Language and Linguistics*. (pp. 275-293). Berlin, New York: Mouton de Gruyter.

Raphael BERTHELE

### Abstract

This study investigates the impact of French corpus planning efforts in two semantic domains, telecommunication and sport. Lists of Anglicisms and their French counterparts pertaining to these two domains are used in a corpus study using the Google Books corpus. A method to explore this corpus diachronically is proposed, and analyses assessing the odds of encountering the ‘official’ French terms over English loanwords are carried out. Despite a certain success of the French competitors in some specific cases, the overall analyses, modelling the odds ratios using mixed effect models, show no sustainable effect of the efforts to impose French alternatives to English neologisms. The results and possible explanations for the statistical patterns are discussed in the light of previous research on Anglicisms and language planning.

Keywords: loanwords, lexical borrowings, linguistic purism, corpus planning, French, Anglicisms, diachronic linguistics

### 1 Introduction<sup>1</sup>

Linguists sometimes forget that the boundaries of their most fundamental categories can be vague. As researchers working on highly diverse and dynamic linguistic ecologies have pointed out, categorizing and counting the ‘languages’ in areas such as the South Pacific is often impossible (Mühlhäusler, 1996). Anthropological linguists have long criticized the use of Western ideologies about ‘language’ for the investigation of language practices in other cultural and geographical areas (Hill, 2002). However, there are also effects of the stereotypical representation of ‘language’ within the Western world, e.g. when the purity of a language is at stake, or more fundamentally when a language needs to be construed as a national, regional or ethnic symbol. Linguists are often central actors in various types of language management (Shohamy, 2006; Spolsky, 2009). Often, the policy measures involved betray stereotypical thinking about what (a) language is, even on the part of ‘experts’. One of the most notorious examples of language management is linguistic purism, when it serves as a form of cultural defense of the modern and post-modern nation:

In the linguistic debate, the specific form of the romantically nationalist position is a concern for the purity of the language. Defending the common language against foreign influences (loan words, basically) is at the same time a defense of the cultural identity of the people. (Geeraerts, 2003, p. 49)

---

<sup>1</sup> Acknowledgements: Many thanks to Jan Vanhove for helping with his statistical expertise. Thanks to Eline Zenner, Elisabeth Dutton and Jan Vanhove for their critical feedback on earlier versions of this text.

As Geeraerts (2003) points out, both rationalist and romantic models of language can provide an ideological basis for linguistic purism. Both models, in my view, relate to instrumental values of language: the former to the instrumental value of language as a tool for communication shared by all citizens of the nation, the latter to the instrumental value of language as a means of identity construction of the ethnically or culturally defined nation. Both ideologies use language as a tool in the construction of the nation, and both perspectives foster arguments in favor of linguistic purism. Purism is needed either to allow smooth communication unhindered by unintelligible foreignisms, or to keep the national symbol clean.

It is common for linguists to condemn linguistic purism as the quixotic struggle of reactionary language planners against the evolutionary processes that cause languages to change. The cognitive sociolinguist, however, might have a more nuanced view on this phenomenon: I see cognitive sociolinguistics as a sub-discipline that tries to understand the interplay of social forces and cognitive processes in shaping the thinking about and the use of languages. From this perspective, linguistic purism is neither *a priori* ridiculous nor ‘un-scientific’. Arguably, there is no fundamental difference between lay ideologies and linguists’ language ideologies: linguists are themselves inevitably ideological when they frame linguistic practices with their categories. In doing so they create languages and varieties such as Makedonian, Singlish, or tussentaal (Geeraerts, Penne, & Vanswegenoven, 2000). Despite many linguists’ self-delusive claims to be ‘descriptivists’ rather than ‘prescriptivists’, their descriptions cannot avoid a normative potential. If, for example, Singlish becomes a recognized category in linguistics, this inevitably leads to the erasure of within-category differences and to the profiling of between-category differences. In this sense, there is only a small step from the cognitive modeling of the chaotic realities of language use to the purist modeling of a language or dialect.

In this paper I assess the impact of a particular attempt at language management – French language purism as instantiated in the Loi Toubon from 1994. The question I try to answer is not whether French or any other linguistic purism is tenable from the linguistic point of view, but rather whether the top-down attempt to steer language change in a particular direction, a form of corpus planning (Haugen, 1987), has any impact whatsoever on language use. In order to do this, I made use of technology that has recently become available to linguists, viz. the Google Books corpus and software packages that facilitate extracting and analyzing data from this corpus. In section 2 I sketch the backdrop of the type of phenomena that I investigate by briefly discussing French corpus planning and purism. Section 3 gives an overview of the corpus and the data extracted from it. In section 4 I provide quantitative analyses of the data. A final discussion (5) focuses on the results and on some critical aspects of the analyses presented.

## 2 Purism and French language management in the 20<sup>th</sup> century

Purism can either focus on the linguistic matter only, or also involve the fight against foreign influences on pattern replication. In this contribution, I use Matras’ (2009) terms *matter* and *pattern replication*. The former refers to the integration of morphological or phonological matter from a ‘foreign’ into a receiving language. The latter refers to the copying of semantics or constructional patterns into the replicating language. Since my main goal is to investigate the impact of ideology and language management on language use, the philological details of the

history of instances of borrowed matter are not relevant. Indeed, in some cases, French matter migrates into English and then back into French (e.g. ‘cash’ or ‘bar’ that were replicated in English based on middle French ‘caisse’ and old French ‘bare’). It is sufficient, for my present purposes, that a linguistic form is perceived as foreign and is thus potentially the target of linguistic purism.

Linguistic purism may also include a fight against pattern replication. For instance, the French linguist Claude Hagège (2011) uses what I refer to as a “Whorfoid ideology” (Berthele, 2014) in his struggle for the status of French as the language of the French (and the world). According to this view, linguistic usage that converges towards patterns of a foreign language also involves convergence towards the thinking patterns of this language (which explains the – in my view misguided – reference to Slobin’s (1996) thinking for speaking approach).

## 2.1 Language management in 20<sup>th</sup> century French legislation

Chansou (1997) provides an overview of the history of French language legislation in the second half of the 20<sup>th</sup> century. There is great continuity across the different legislative actions discussed: The main goal is to promote the status of French in the face of the increasing importance of English as a global language. However, different types of action and also different aspects of language are in focus at different times. As shown in Chansou (1997), the policy implemented under Presidents de Gaulle and Pompidou is a “dirigist” attempt to impose French instead of English or other foreign terms in the educational context, inspired by treatises such as Etiemble’s “Parlez-vous français?” (1964). A decree from 1972, entitled “enrichissement de la langue française”, spells out the mechanism of corpus planning in the domain of terminology (“proposer les termes nécessaires soit pour désigner une réalité nouvelle soit pour remplacer des emprunts indésirables aux langues étrangères”; cf. Chansou 1997: 26). Worth mentioning are some rather comprehensive and constraining predecessors of the currently applicable law, e.g. a text submitted to Parliament in which not only borrowings (in the sense of matter replication) are prohibited, but loan translations, i.e. the replication of ‘foreign’ patterns, as well:

« Sont prohibés le recours à des termes étrangers ou dérivés de termes étrangers ou l'utilisation de formes de langue calquées sur un modèle étranger. » Proposition de loi n° 306 relative à la défense de la langue française présentée par Pierre Bas le 10 mai 1973, article premier ; cf. (Chansou, 1997, p. 31)

For the sake of brevity I will not discuss the different stages that led to the current central legal regulations, most prominently instantiated by the Loi Toubon that was submitted to the French Senate in 1993. As a general tendency, the rather strict and dirigistic view of language management in earlier legal texts gradually shifted towards what is now called the “enrichment of the French language” via suggestions from a terminology committee (“Commission générale de terminologie et de néologie”), as stated in a decree from 1996. The terminological suggestions are obligatory in documents produced by government agencies and services, and in the “Journal officiel de la République française”, the daily bulletin of the French government. The French neologisms proposed by this committee are published in different channels, among others in thematically grouped brochures (see below), two of which were used for the analyses in this contribution.

Maybe due to the critical view held by most scholars regarding linguistic purism, there is only rather scarce work on the actual impact of these terminological recommendations. Traditionally, as Humbley (2008, p. 85) discusses, linguists think that such attempts are both wrong and bound to fail. At least two scholars, however, acknowledge that in particular in the domain of IT terminology, there is a detectable impact of corpus planning efforts on language uses (Depecker, 2001; Kaltz, 1988). From the methodological point of view, the suggestions made by Humbley (2008) are not really satisfying for linguists interested in usage-based approaches: Comparing dictionary entries across time is certainly a first attempt, but cannot provide reliable evidence for actual patterns of language use. And simply counting hits in a search engine lacks the necessary diachronic dimension. As argued by Zenner, Speelman and Geeraerts (2012, 2014), counting occurrences of the loanword only is insufficient, also because this procedure does not take into account the topic specificity. As Zenner et al. argue, only a method that includes receptor language alternatives to the Anglicisms will provide valid insights into the diffusion of borrowed words or their terminological alternatives. In the remainder of this paper I will propose a method of how the possible impact of these recommendations could be measured.

### 3 Method and Data

This section describes the data used to answer the research questions formulated above. First, I describe how I identified the lexical candidates based on the official terminological recommendations. Second, I describe briefly how the French\_2012 version of the Google Books corpus can be used to extract occurrences of terminological pairs (a potential English borrowing and its French correspondence) that are in competition.

#### 3.1 Bulletin officiel

The language management process described in section 2 leads to a great number of terminological recommendations in different domains. To keep the amount of work and data within reasonable bounds, I selected two domains for further analysis: sport and telecommunication. Both domains are highly international and thus potential fields of contact with and borrowing of international, i.e. English, terminology. The two brochures (Premier ministre & Commission générale de terminologie et de néologie, 2009, 2011) are freely available from the webserver of the French government.<sup>2</sup>

These brochures contain suggested French equivalents to international terms, as illustrated in Figure 1.

---

<sup>2</sup> <http://www.culture.fr/Ressources/FranceTerme/Vous-pouvez-le-dire-en-francais> [consulted on 26.2.2015]

**parrainage**, n.m.

◆ **Domaine** : Économie et gestion d'entreprise. ◆ **Définition** : Soutien matériel apporté à une manifestation, à une personne, à un produit ou à une organisation en vue d'en retirer un bénéfice direct. ◆ **Note** : Les opérations de parrainage sont destinées à promouvoir l'image du parraineur et comportent l'indication de son nom ou de sa marque. ◆ **Voir aussi** : mécénat, patronage. ◆ **Équivalent étranger** : sponsoring, sponsorship.

Source : *Journal officiel* du 22 septembre 2000.

Figure 1: An entry in the sports brochure

From these entries, the foreign word (“équivalent étranger”) and the French word were extracted. These lexical pairs were written into an item table that also contains the year of norm-imposition (2000 for ‘parrainage’) as well as part-of-speech information.

If, as in the case shown in Figure 1, the entry contained several variants (sometimes both on the French on the “foreign” side), I selected the most frequent words. If two spellings were possible in either English or French, the most frequent one was chosen (<freestyle> vs. <free-style>). The adjectives in the list were discarded, since there were only few of them. Only verbs and nouns were considered for further analyses. Extremely polysemous items were discarded as well: French ‘dehors’ and English ‘out’ are frequent adverbs and do not only refer to the footballing concept ‘out’; their presence in the list of search terms would hardly be informative for the investigation of usage patterns of the sports term. Similar problems arise with terms such as French ‘lièvre’ for the ‘pace-maker’ in athletics. If abbreviations were more frequently used than their spelled out forms (‘ADSL’ instead of ‘asymmetric digital subscriber line’), the former were used as search terms. Identical terms (‘modem’ both in English and French) in both languages were deleted from the list of analyzed terms.

### 3.2 Google Books corpus

In order to investigate use patterns of the terms listed by the Academy committee, absolute occurrences in the French Google Books corpus were extracted (see Michel et al., 2011 for further information on the Google Books corpus). This was done using the ngramr package (Carmody, 2013) and a script in the statistical programming language R that automatically extracted the frequencies of pairs of terms across the different corpora per year. The script is freely accessible online at [http://figshare.com/articles/Googling\\_Toubon/1321270](http://figshare.com/articles/Googling_Toubon/1321270). The period from 1950 to 2008 was selected (no newer data are currently available). Figure 2 shows the size of the French Google Books corpus (version 2012) in millions of words.

Preliminary versions of the search revealed that the ‘French’ corpus also contains texts in English. The search for English search terms thus also potentially yields hits within English books, texts or longer citations inside French texts. For the nouns, this unwanted effect was minimized by adding the “\_DET\_” tag before both French and English search terms. Test searches proved that this leads to hits that only contain French determiners (*le, la, les, un, une, des, l’*), even for the English search terms. For the verbs, unfortunately, no such procedure that leads to a purer proportion of French language hits could be found, since adding the POS tag “\_VERB\_” also yields hits from English text passages within the French corpus.

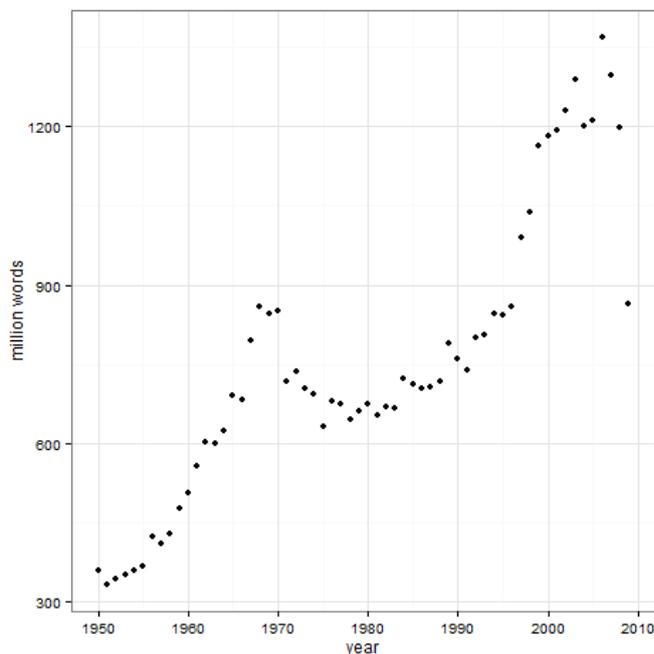


Figure 2: The Google Books French 2012 corpus across the time period under investigation

The search for corpus occurrences for the terms of the two domains produced a large number of empty results, both in French and in English. For example, neither the English loan word ‘aquabike’ nor the French term ‘aquacycle’ produced any hits. In 173 cases, the English term is never found in the corpus (e.g. ‘announcer’), only the French (e.g. ‘annonceur’). One reason for

this could be that the foreign terms in the brochure are not exactly the terms used in French. At least in one case, ‘cache memory vs. antémémoire’, the term actually used, at least as far as my personal observation goes, is ‘mémoire cache’, i.e. a French noun modified by an English loanword. In this case, I modified the ‘foreign’ search term in order to reflect a usage pattern involving the target loan word that can actually be observed in French. Lastly, in other cases, no hits were found for the French term (‘canyonisme’), but only for the English (‘canyoning’).

As Table 1 shows, a majority of the terms subject to regulation by the terminology committee do not show up at all in the Google Books corpus. It is thus safe to say that they have only rather marginal usage frequencies, at least in written publications. Only 30 sports terms and 151 telecom terms produced any hits in the period under investigation that can be used for the analysis below.

pairs of noun expressions:	Sports	telecommunication
total pairs in search list	206	595
neither French nor English ngrams produced	111	239
no French ngrams produced	37	60
no English ngrams produced	28	145
total usable	30 (28 N, 2 V)	151 (130 N, 21 V)

Table 1: Number of search terms per domain

#### 4 Analyses

The data collected as described above were analyzed descriptively and inferentially. In both cases, the overarching questions were (a) whether there are changes over time in the probability of use of the French vs the English variant of the pairs, and (b) whether a change in relative frequency can be observed after norming in the *bulletin officiel* took place.

##### 4.1 Descriptives

In the figures below, the relative frequencies in each pair of terms are displayed in odds ratios of the choice of the French variant.<sup>3</sup> The dependent variable thus stands for the odds that the French term was chosen over the English competitor. Figure 3 plots the development of the odds for two selected pairs of terms across the period investigated.

---

<sup>3</sup> odds ratio: (count of hits for French term + 0.5) / (count of hits for English term + 0.5)

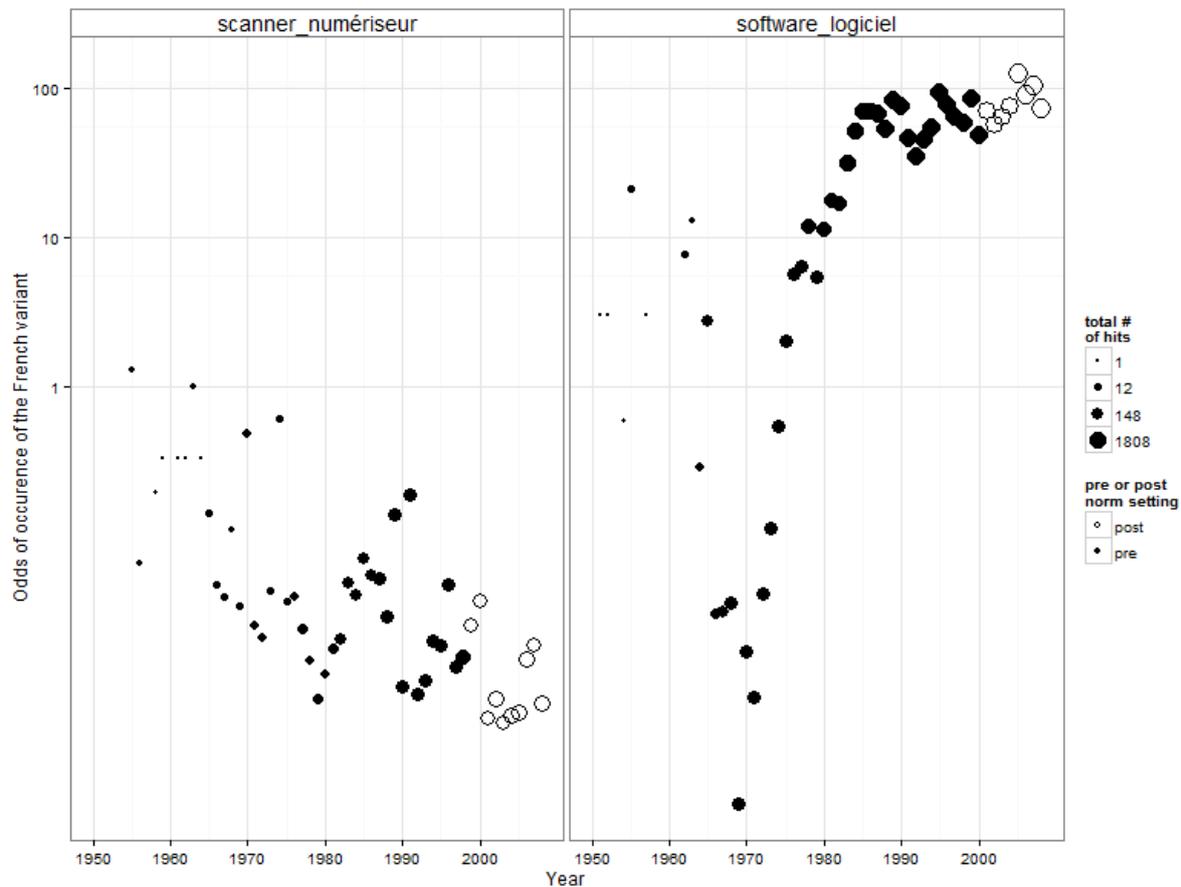


Figure 3: Odds for the choice of the French variants of two selected items from the telecom domain across time. The hits for the French and English competitors are expressed in the odds of encountering the French word. As an example, a value of 100 indicates that for each hit of ‘software’ there are 100 hits of ‘logiciel’. Larger point sizes indicate a larger sum of hits for both terms combined. The solid points refer to corpus frequencies before the terminology committee has issued a language norm for the given term. The y-axis is log-transformed.

The plot for the pair ‘software vs. logiciel’ indicates that, at the beginning of the period investigated, only a few occurrences are found in the corpus. In many years until the mid-1970s, the odds of the French variant are below or around 1, which means that the English word was as likely or even more likely to be used than the French term. At the end of the period, the odds of encountering the French term are about 125 times larger than of encountering the English term (odds of 125:1). It is probably examples such as this one that led some scholars to the conclusion that corpus planning, at least in the IT domain, was a “big success story” (Humbley, 2008, p. 93). Indeed, if all terms in the sample showed this development, one would have evidence for a shift in preference towards the French terms. However, it is quite obvious that not all items do, e.g. the other item in the figure (‘scanner vs. numériseur’), shows the opposite pattern of change over time, with odds around 0.006 for the French term, i.e. the French term is selected 0.006 times for each English term selected. It is more intuitive to express the odds of the occurrence of the English term in this case, which would be around 175 for each French term.

As the two items highlighted show, different types of changes in frequency before and after the year norm-setting took place can be observed. Figure 4 gives the same account for two items from the domain of sports.

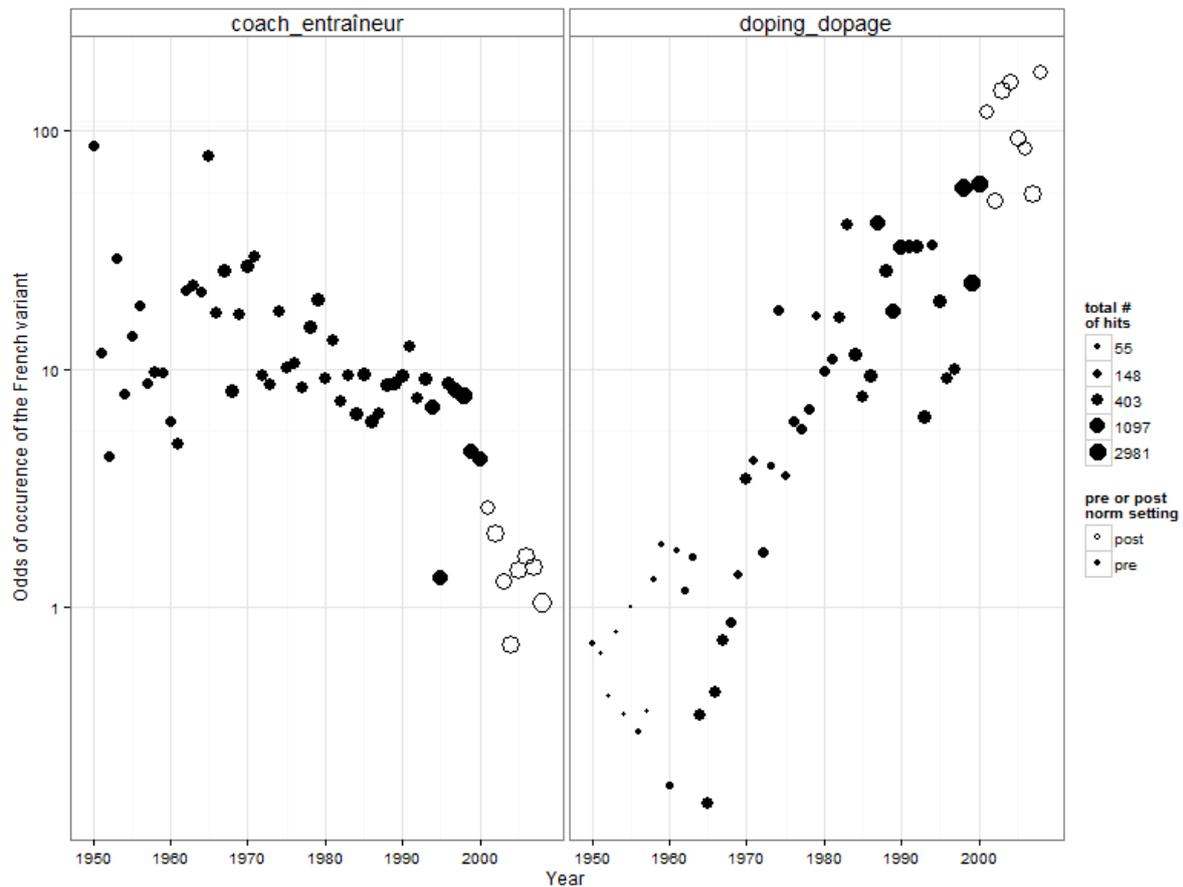


Figure 4: Odds for the choice of the French variants of two selected items from the telecom domain across time. The solid points refer to corpus frequencies before the terminology committee has issued a language norm for the given term.

As the two figures illustrate, the odds of the choice of the French variant can change in different ways across time. From the visual inspection of each item pair's plots (not shown here), three different types of development can be observed.

- 1) General trend towards French
- 2) General trend towards English
- 3) No trend, odds do not seem to change

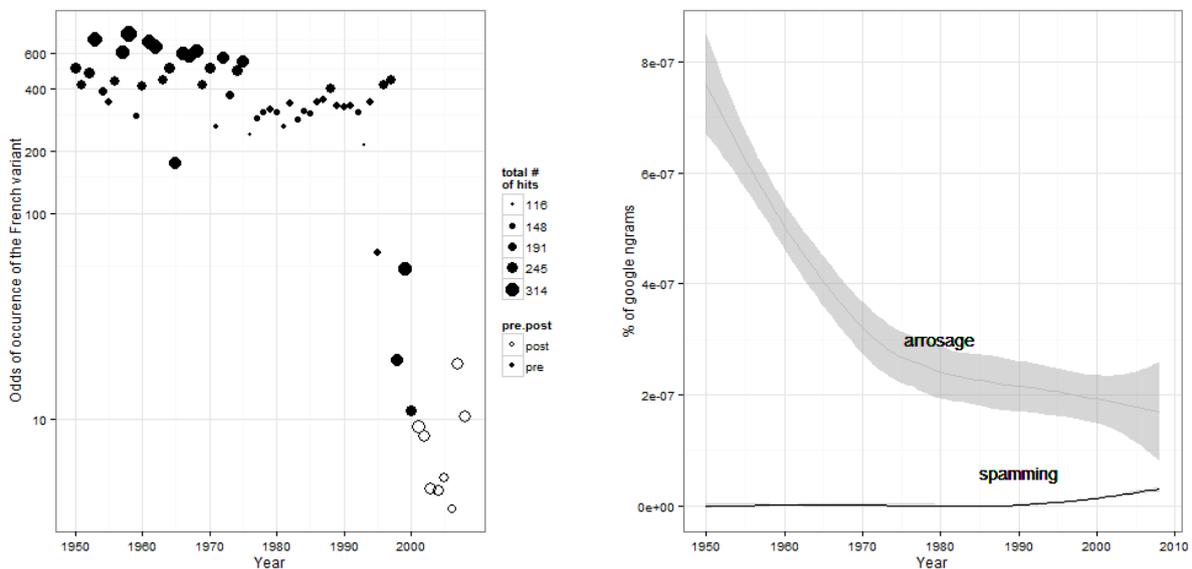
The example 'doping vs. dopage' in Figure 4 is an example of (1). Examples of other, similarly patterning terms are 'disk vs. disquette' or 'click vs. cliquer'. The example 'scanner vs. numériseur' in Figure 3 is an example of (2). Other, similarly patterning terms are 'hub vs. concentrateur' or 'ADSL vs. RNA'. In these latter cases, as Figure 5 below illustrates, the trend

towards English is often the consequence of a pre-existing French word used to refer to a non-telecom entity.

Figure 5 thus shows that the French word was used without much competition from English ‘spamming’. ‘Spamming’ is almost absent from the corpus until the mid-nineties. From the moment that the practice of spamming was first observed,<sup>4</sup> the English term becomes increasingly frequent relative to the French competitor. For readers who are used to the Google ngram-viewer, I added ‘ngram-style’ plots with smoothers based on the relative frequencies on the right hand side of plots 5–7.

Since the French terms suggested by the committee are often polysemous, the numbers produced by the procedure applied do not necessarily directly reflect the competition of the two terms as a means to convey the particular sports or telecom semantics. Obviously, spam via email was not a major problem before about the year 2000 and thus the loanword was highly unlikely to show up in French texts anyway; at the same time, the term ‘arrosage’ is used in its literal sense, i.e. to refer to ‘watering’ (flowers, etc.). The change in relative occurrences across time is thus probably not due to some policy change, but simply to the fact that an English term was borrowed to refer to an emerging phenomenon. The size of the data points in Figure 5 between the years 1975 and 1995 shows that the number of hits was relatively low before the phenomenon of spam messages arose around the year 2000.

One could argue that the pair of terms should not be included also due to the polysemy of the French term. Only corpus searches taking into account more neighbors and probably also involving manual selection of search results could separate the wanted from the unwanted tokens. Despite this presence of unwanted tokens in the data, there is still the question whether it is nevertheless possible to detect changes in odds from the onset of spam practice on. We will give a tentative answer to this question below in section 4.2.



<sup>4</sup> According to Templeton <http://www.templetons.com/brad/spamterm.html>, the message “Global Alert for All: Jesus is Coming Soon.” was sent to all USENET members on January 18 1994.

Figure 5: The usage frequency of ‘arrosage’ vs. ‘spamming’ across the observation period.

As shown in Figure 5, there are item pairs that shift across time towards the use of English. Others, as already illustrated in Figure 3 and Figure 4, shift towards the use of French: The verb ‘cliquer’ (a Francized verb based on English ‘to click’) is an example of change towards the use of the French word (Figure 6).

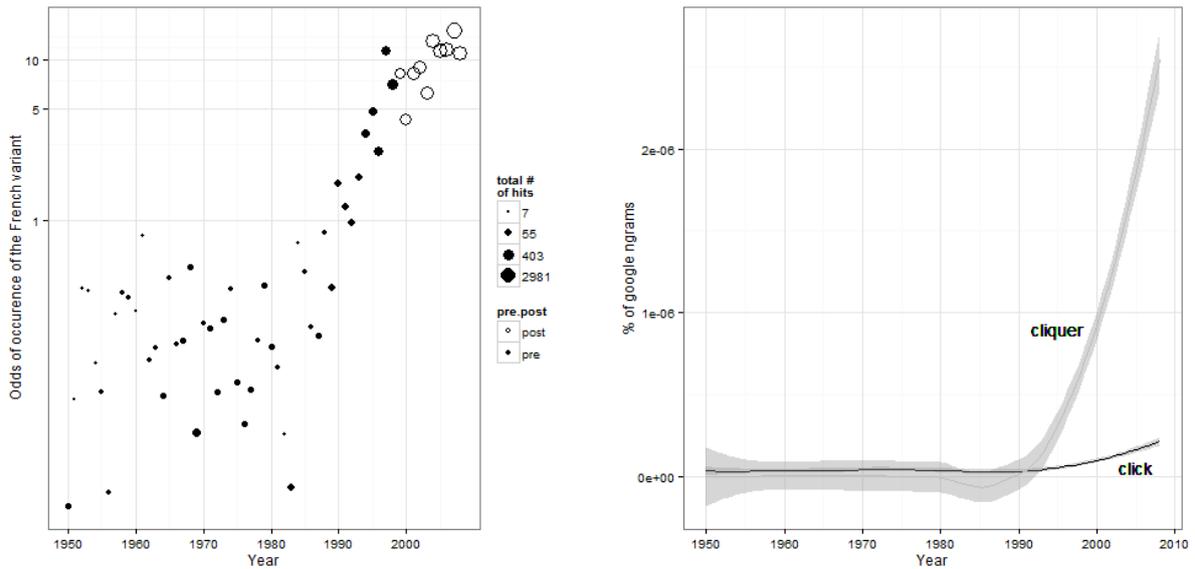


Figure 6: Odds and frequencies of the verb pair ‘click vs. cliquer’

A third group of items does not display any striking changes of odds at all. As an example, the pair of verbs ‘sponsor vs. parrainer’ does not shift into either direction, as shown in Figure 7.

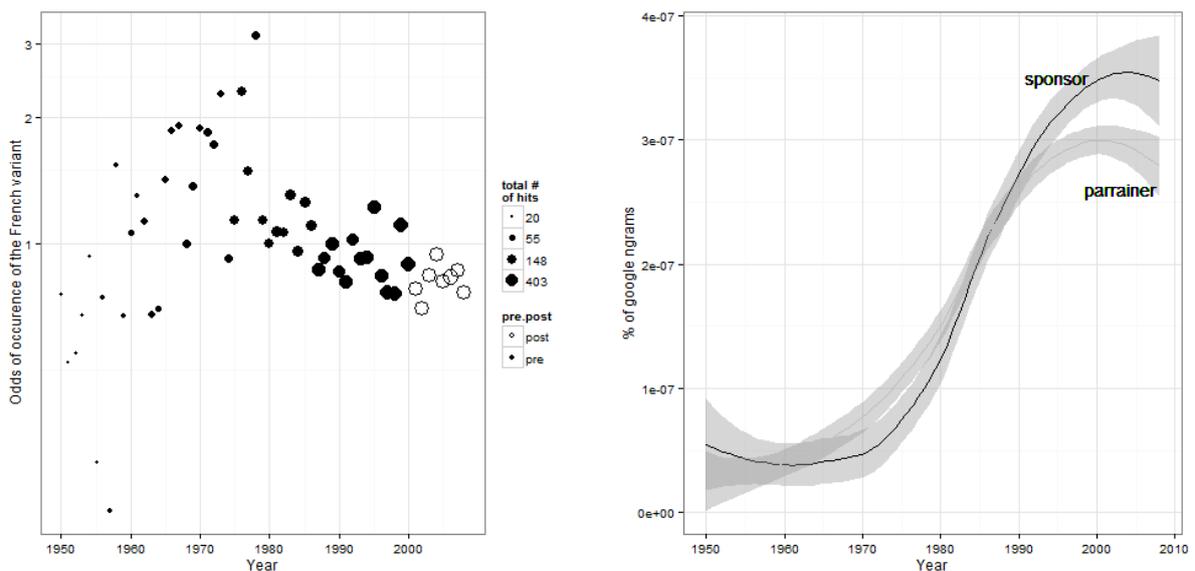


Figure 7: Odds and frequencies of the verb pair ‘sponsor’ vs. ‘parrainer’

To sum up this descriptive analysis of the data and of some selected examples, we can conclude that no clear tendency can be read out of (or should be read into) the data. Some item pairs display a change in the direction of the language planning process intended by the French legislator, others show no change at all, and others even change in the opposite direction. None of the items discussed in more detail above suggest that there is a strong change after the committee issued a recommendation with respect to the use of a particular French variant to the English term. As in Figure 5 and Figure 6, the tendency either towards English or towards French was already under way before, or there is simply no tendency as in Figure 7. We will come back in the final section to the question of the existence and the direction of a causal link between tendencies in use patterns and imposed language norms by the committee in charge of the enrichment of the French language.

#### 4.2 Statistical modeling

In order to assess the impact of the language enrichment policy implemented in France, an attempt to model the development using inferential statistics seems appropriate. Such an attempt is made in this section using logistic linear mixed effects models.

The dependent variable is the odds of encountering a French term given the absolute frequencies of the French and the English competitors in the Google Books corpus. This is the variable selected for the y-axis in all corpus based figures above. The first predictor modeled is time. Change over time in general could hint at an effect of the policy calling for the use of French linguistic matter, as part of the general goal of official French language policy. The question asked, therefore, is whether there is a general tendency increasing the odds of the choice of French in the data. Moreover, since we know exactly in which year the *bulletin officiel* issued a terminological recommendation for each foreign term, we can also ask the additional question whether there are detectable changes, compared to the *terminus ante quem*, in linguistic choices in the years following the issue.

Although one might be tempted to let these two factors interact, it is problematic to hypothesize that the norm-imposing process had a different influence before it even existed compared to afterwards – which is what an interaction of between ‘time’ and ‘before vs after norming’ would be modeling. Thus, it seems more appropriate to include the years since norm-imposition for each item pair as a second main effect into the model (0 when before norm-setting). This modeling approach is akin to piecewise regression (or segmented regression). If the parameter of this main effect turns out to be significant – even after an overall effect of time is taken into consideration–, this would count as evidence for a measurable impact of the norm-imposition.

In addition to these two main effects, all item pairs are modeled using random intercepts and random slopes for the two predictor variables, since, as we have seen above, item pairs tend to show different diachronic patterns, and the model should account for this variation. The models were fitted in R (version 3.1.2) with the lme4 package (version 1.1-7; Bates, Maechler, & Bolker, 2014).

There is a statistically significant tendency towards the choice of the English variant across the time period investigated (beta = -0.194, standard error = 0.059,  $p < 0.001$ ). The low estimate and high p values of the “years since norming” variable shows that there is no detectable change in the slope of the regression line after the norming of the terminology committee (beta= 0.080, standard error= 0.153,  $p=0.603$ ). A likelihood ratio test was used to compare this model to the simpler model that only contains one fixed effect (year of publication) and discards the effect of norming. This test shows that the second main effect can be dropped (Chi-Square (1)=0.27,  $p=0.60$ ). Thus, the additional parameter modelling the impact of terminological norming does not contribute in any noteworthy way to the explanation of the patterns in the data.

These estimates suggest that there is a general tendency towards more use of the English terms, since the negative value of the estimate stands for a declining slope of the regression line. The additional fixed effect representing the potential impact of the norming intervention (*years since norming*) does not yield any noticeable change in slope.

## 5 Discussion and Conclusion

The analysis of the data overall suggests that the drive for linguistic purism which expresses itself, among other things, in the French language legislation discussed in section 2, does not seem to allow a reversal of the tendency of increasing use of English loan words. In a nutshell, the model confirms the general tendency towards the use of the English terms in the two domains, and there is no evidence for any impact of the norming intervention.

The discussion of these results comes with a caveat: The statistical procedure applied presupposes that the development is linear. However, as some of the figures shown above suggest, changes may well be non-linear. This could be taken into account by fitting non-linear mixed models (Wood, 2012). Fortunately, fitting such models and discussing their output are beyond the space limitations of this article – since, unfortunately, they would also require statistical skills that lie way beyond those of its author.

Another point worth considering is that the time span between the committee’s act of norming and the observable language uses is possibly not long enough. There might be more time required before the new norms sink into authors’ language use. Moreover, texts published in a given year after the norming took place may well have been written *before* this norming, which could be a source of error obscuring possible effects of the norming. The arbitrarily chosen domains of sports and telecom may be domains where norming is particularly unsuccessful; in other domains the committee’s recommendations might bear more fruit.

The absence of any measurable impact of the committee’s attempts to influence language use may not surprise scholars studying language change. Changing – potentially – millions of language users’ lexical preferences means changing the sociolinguistic constraints that shape the acquisition and use of those variants. These constraints or forces are often strong and certainly hard to overcome with official policy measures. The institutional banning of loanwords stemming from a language that incarnates economic, political, cultural and scientific strength requires changing the communicative maxims that motivate the speakers to prefer those borrowings to ‘native’ words of the language they use (see Croft, 2000; Keller, 1994 for theoretical models of language change ). Moreover, from the legal perspective, imposing the selection of a language or of particular terms, at least in western societies, is possible only in highly regulated domains such

as text production in the central administration. Imposing terms and languages in other, less controlled domains, goes against the fundamental rights and freedoms cherished in Western societies. In some exceptional cases, however, it might be possible for a state agency to have an impact on language use, as the case of institutionalized Icelandic language purism (Árnason & Helgadóttir, 1991) seems to suggest. What exactly the requirements for the success of such language policy are remains to be investigated.

The present contribution does not lend any support to the claim that the French effort to control and impose linguistic matter in sports and telecommunication terminology is crowned by success. This ties in with the conclusions of another corpus-based study on the productivity of loanwords that showed that English loans in French are distinctively productive (Chesley, 2010). When I started my investigation, I was surprised that it was difficult to find corpus-based analyses of the actual effects the French language legislation. In retrospect I conclude that there are two reasons that this might not be so surprising. First, it may well be that the authors of the terminological recommendations themselves do not believe in the potential effectiveness of their actions. Second, the real function of French language purism and its institutional instantiations might be one that lies well beyond their potential effects on users of French. They stand as a symbol, as an emblem for a cultural model of French as a language that does not need to import terms from other languages. In this sense, these policy measures are performative components of the cultural models of language described in Geeraerts (2003), i.e. their mere existence is what really counts.

## 6 References

- Árnason, K., & Helgadóttir, S. (1991). Terminology and Icelandic Language Policy *Behovet och nyttan av terminologiskt arbete på 90-talet. Nordterm-symposium.* (pp. 7-21): Nordterm 5.
- Bates, D. M., Maechler, M., & Bolker, B. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. <http://cran.r-project.org/package=lme4>
- Berthele, R. (2014). Biestmilch, Schafspferche und Schamanen: Überlegungen zur Verwendung whorfoiden Gedankenguts im Diskurs über sprachliche Diversität. *Zeitschrift für Semiotik*, 35(1-2), 85-107.
- Carmody, S. (2013). ngramr. Retrieve and plot Google n-gram data. Version 1.4.5.
- Chansou, M. (1997). Les politiques de la langue et la législation linguistique en France (1966-1994). *Mots*, 23-35.
- Chesley, P. (2010). Lexical borrowings in French: Anglicisms as a separate phenomenon. *Journal of French Language Studies*, 20(03), 231-251. doi: doi:10.1017/S0959269510000049
- Croft, W. (2000). *Explaining Language Change*. Harlow: Longman.
- Depecker, L. (2001). *L'invention de la langue : le choix des mots nouveaux*. Paris: A. Colin : Larousse.
- Etiemble. (1964). *Parlez-vous franais?* Paris: Gallimard.
- Geeraerts, D. (2003). Cultural models of linguistic standardization. In R. Dirven & M. Pütz (Eds.), *Cognitive Models in Language and Thought. Ideology, Metaphors and Meanings* (pp. 25-68). Berlin/New York: de Gruyter.

- Geeraerts, D., Penne, A., & Vanswegenoven, V. (2000). Thuis-taal en Familie-taal. Taalgebruik in Vlaamse soaps. In S. Gillis, J. Nuyts & J. Taeldeman (Eds.), *In Met taal om de tuin geleid. Opstellen voor Georges De Schutter* (pp. 161-170). Antwerpen: UIA.
- Hagège, C. (2011). *Contre la pensée unique*. Paris: O. Jacob.
- Haugen, E. (1987). Language Planning. In U. Ammon, N. Dittmar & K. Mattheier (Eds.), *Sociolinguistics. An international Handbook* (pp. 626-637). Berlin, New York: De Gruyter.
- Hill, J. H. (2002). "Expert Rhetorics" in Advocacy for Endangered Languages: Who Is Listening, and What Do They Hear? *Journal of Linguistic Anthropology*, 12(2), 119-133.
- Humbley, J. (2008). How to determine the success of French language policy on Anglicisms - some methodological considerations. In R. Fischer & H. Placzewska (Eds.), *Anglicisms in Europe. Linguistic Diversity in a Global Context* (pp. 88-105). Cambridge: Cambridge Scholars Publishing.
- Kaltz, B. (1988). EDV-Sprache. Zu einige Wortbildungstendenzen im Deutschen und Französischen. *Muttersprache*, 98, 38-49.
- Keller, R. (1994). *On language change: the invisible hand in language*. London, New York: Routledge.
- Matras, Y. (2009). *Language contact*. Cambridge: Cambridge University Press.
- Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Team, T. G. B., . . . Aiden, E. L. (2011). Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*, 331(6014), 176-182. doi: 10.1126/science.1199644
- Mühlhäusler, P. (1996). *Linguistic ecology : language change and linguistic imperialism in the Pacific region*. London: Routledge.
- Premier ministre, & Commission générale de terminologie et de néologie. (2009). Vocabulaire des techniques de l'information et de la communication (TIC). Termes, expressions et définitions publiés au Journal officiel. *Enrichissement de la langue française*.
- Premier ministre, & Commission générale de terminologie et de néologie. (2011). Vocabulaire des sports. Termes, expressions et définitions publiés au Journal officiel. *Enrichissement de la langue française*.
- Shohamy, E. (2006). Manipulating languages. In E. Shohamy (Ed.), *Language policy : hidden agendas and new approaches* (pp. 22-44). London ; New York: Routledge.
- Slobin, D. I. (1996). From 'thought and language' to 'thinking for speaking'. In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking Linguistic Relativity* (pp. 70-96). Cambridge: Cambridge University Press.
- Spolsky, B. (2009). *Language management*. Cambridge ; New York: Cambridge University Press.
- Wood, S. N. (2012). mgcv: Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation. R package, version 1.7-22. . <http://cran.r-project.org/package=mgcv>
- Zenner, E., Speelman, D., & Geeraerts, D. (2012). Cognitive Sociolinguistics meets loanword research: Measuring variation in the success of anglicisms in Dutch. *Cognitive Linguistics*, 23(4), 749-792. doi: DOI 10.1515/cog-2012-0023
- Zenner, E., Speelman, D., & Geeraerts, D. (2014). Core vocabulary, borrowability and entrenchment A usage-based onomasiological approach. *Diachronica*, 31(1), 74-105. doi: DOI 10.1075/dia.31.1.03zen