

The World Wide Web as a Linguistic Corpus

Russon Wooldridge
University of Toronto
wulfric@chass.utoronto.ca

Abstract

The largest electronic text database in existence, the World Wide Web, makes possible, for the first time and without any outlay of costly set-up or the expensive development of project-specific tools, the observation of current usage in a number of different languages. The Web as database has also the enormous advantage over other language corpora of behaving like natural language: it is dynamic, unceasingly renewing itself and thus offering snapshots of the present state of the language, with its proportion of new, established and aging usage. The paper discusses the observation, since 1998, of French, including derivation, polysemy, homophony, syntagmatic variation and the genesis of new concepts and their linguistic naming. The language corpora observed are those of micro-systems -- particular words, word families and lexical constructions -- and not the macro-system of the language as a whole. The tools used in the observation are keywords and search engines, in particular Google.

KEYWORDS: World Wide Web, corpus linguistics, WWW as database, polysemy, homophony, syntagmatic variation, neology.

Introduction

The largest electronic text database in existence, the World Wide Web, makes possible, for the first time and without any outlay of costly set-up or the expensive development of project-specific tools, the observation of current usage in a number of different languages. The Web as database has also the enormous advantage over other language corpora of behaving like natural language: it is dynamic, unceasingly renewing itself and thus offering snapshots of the present state of the language, with its proportion of new, established and aging usage.

The paper discusses the observation, since 1998, of French, including derivation, polysemy, homophony, syntagmatic variation and the genesis of new concepts and their linguistic naming. The language corpora observed are those of micro-systems -- particular words, word families

and lexical constructions -- and not the macro-system of the language as a whole. The tools used in the observation are keywords and search engines, in particular Google; the study takes advantage of the search engine's highlighted display of keywords in result summaries, in cache copies of online HTML documents and in HTML copies of online documents in PDF or DOC format.

1. The Norman word *douet*

On the verges of the winding lanes in the Pays d'Auge in Normandy can be seen signs announcing that one is following the "Route des Douets". Before going to the library to consult linguistic atlases or regional glossaries one can search for the word *douet* in the French-language pages of the Web:

"La route des « douets » (ruisseaux en patois Normand) passe par ici et alimente un très ancien lavoir" (Page on Saint-Hymer, Pays d'Auge)

"La Route des Douets. Elle vous fera connaître les eaux-vives du Pays d'Auge, ces « douets », petits ruisseaux allant se jeter dans la Touques, et quelques rivières telles que l'Yvie et la calonne." (Page on Villers-sur-Mer, Pays d'Auge)

"Le bourg est traversé par un petit cours d'eau, le Doig (appelé aussi le Beuvronnais), qu'il suffit de prononcer à la normande pour retrouver ce qui est plus probablement son sens réel: le Douet, signifiant ruisseau en augeron (patois du Pays d'Auge)." (Page on Beuvron-en-Auge)

Thus a few seconds of querying yield both linguistic and pragmatic information, the latter more difficult to find in atlases or glossaries.

2. Word Families and derivation

2.1. *Enfirouaper* and its derivatives

The Canadian verb *enfirouaper* receives the following treatment in the

Dictionnaire québécois d'aujourd'hui (DicoRobert, 1993):

“Fam. embrouiller, emberlificoter qqn pour le tromper, le rouler. => **duper, enjôler, entortiller, séduire**. *Se faire enfirouaper par un vendeur*. => **berner, jouer** ; fam. **pogner** ; très fam. **fourrer**.”

This is a semantic and onomasiological portrait with just one example of usage. What however is the real usage of the word ? In 1998 and 1999 the offerings of the Web included the following:

“Comme un politicien, le génie cherchait un moyen
D’m’enfirouaper, c’est ben certain, c’t’un beau menteur
Hey!” (Extract from *Le génie dans la bouteille*, poem)

“A quoi servent les gestes?
A arracher, à séduire,
A enfirouaper, à conquérir;
Après, qu’est-ce qu’il reste?” (Extract from *A quoi servent les mots?*, poem)

“Le romancier Yves Beauchemin (*Le Matou*) qui lui a dédié son roman, *L’Enfirouapé* résume ainsi la carrière de ce salonnier montréalais.” (Page on Henri Tranquille)

“Je suis la dernière arrivée. Denyse est sûrement une philanthrope ou plutôt une « félinthrope », une sensible qui se laisse enfirouaper. Autrement, elle m’aurait amené à la SPCA.” (Extract from *Denyse Mageau vue par ses trois chats*)

“Mais, comme toujours, les francophones au Québec n’ont été surpassés que par leurs congénères d’Ottawa, politiciens et fonctionnaires. Et le Canada anglais s’est fait joliment « enfirouaper ».” (Expression of political opinion)

“M. Gauthrin: M. le Président, si vous le permettez, j’ai présenté une motion sans préavis. Le gouvernement dit

oui ou non. Maintenant, qu'ils essaient d'enfirouâper ça pour toutes les raisons pour lesquelles ils disent non, moi, je n'en sais rien." (Debate of the Assemblée nationale)

"Pas plus que la veille quand il s'était rué sur Eddie Irvine, pour lui dire son fait, un Villeneuve frondeur ne s'est laissé intimider, dans un affrontement décisif, avec l'Allemand qu'on disait tellement fort qu'il allait l'enfirouaper. Une saison de fou pour Jacques qui a démontré qu'il avait toute une carapace." (*Le Soleil*, newspaper)

The word is onomatopoeic, used in verse and prose, uttered by human and animal alike, political, a fighting word, found in a variety of language registers.

The search engine AltaVista, unlike Google, allows one to discover lexical forms through the use of an end-of-word wild card (asterisk). The query **enfirou*** found, among other forms, the derivatives **enfirouapage** and **enfirouapeux**:

"Considérant les commentaires de tout un chacun depuis que la Cour suprême a émis son avis, il semblerait que peu de gens l'aient scruté. D'autres l'ont lu, ou parcouru, mais n'en retiennent publiquement que les éléments qui leur conviennent un peu comme s'il s'agissait d'un bar à salade. On appelle cela de l'enfirouapage." (Site *Canada Ensemble*)

"Le ministre de l'enfirouâpage systématique étale son intransigeance crasse. Aujourd'hui, il fait dire aux juges de la Cour suprême ce qu'ils n'ont jamais dit." (Site *Vigile* « *Je me souviens* »)

"Mais voilà, les politiciens enfirouapeux et malhonnêtes font [sic] légions [sic]." (*Page de Luc Richard*)

One notes the high degree of semantic (shades of meaning, registers) and morphological (derivatives) productivity of the verb *enfirouaper*.

2.2. The family of the verb *scribouiller*

We shall take as our starting point a well-known sentence uttered by the Général de Gaulle:

“Et quant au reste, tout ce qui grouille, grenouille, scribouille, n’a pas de conséquence historique dans ces grandes circonstances — pas plus qu’il n’en eut jamais dans d’autres.” (Taken from a speech given in Montreal in 1967)

The *Trésor de la langue française* (TLF) (the word is absent from the majority of less extensive dictionaries) has this to say:

“**SCRIBOULLER**, verbe intrans. *Fam., péj., vieilli* A. Être employé aux écritures. (Dict. XXe s.). **B.** Écrire sans soin ou sans talent. (Dict. XXe s.)” (*Trésor de la langue française*)

Of interest is the second meaning of *scribouiller*, “écrire sans soin ou sans talent”. To this can be added the definitions given by the TLF for *scribouilleur* “Celui qui scribouille; auteur, écrivain sans talent” and *scribouillage* “Action de scribouiller; résultat de cette action”.

The Web contains de Gaulle’s sentence, of course, but especially (hundreds of occurrences) the verb *scribouiller* used in a colloquial, attenuative register in online chats. For example:

“Quand je pense qu’à l’heure où je scribouille l’auteur du polar fait bronzette sur la passerelle d’un cargo mexicanos qui transporte 5000 tonnes de chiquitas pour le marché européen...” (List polaroid@galeode.fr)

We shall look now at the noun **scribouilleur**, or rather *scribouilleur*, *scribouilleuse*, and *scribouilleux*. In June 2003, the Web contained -- i.e. Google found -- 264 occurrences of *scribouilleur(s)*, 987 of *scribouilleuse(s)* and 7 of *scribouilleux*.

A *scribouilleur* is, as the TLF states, “celui qui scribouille; auteur, écrivain sans talent”:

“Henning Mankell est bien plus qu’un simple scribouilleur de petits polars de divertissement; c’est un fin observateur d’une société en perdition, un humaniste dans l’âme qui cherche à disséquer et tente de comprendre ce monde étrange et malade dans lequel il vit, et nous tous avec lui.” (Marie-Claude Mirandette dans *Le Devoir* en ligne)

On the other hand, a *scribouilleux* is rather someone who *scribouille* according to the norms of online chat:

“Scribe: Les forums de la FPJQ? Ben non, à peu près tout ce qui compte de scribouilleux dans la province ont déjà reçu ta prose et ont installé des bozos filters en conséquence.” (Forum *Le Code*) Note that it is the *scribe* who moderates this type of forum.

As for *la scribouilleuse* (the feminine is over three times more frequent than the two forms of the masculine combined), she appears to owe her success to the name of an online journal:

“Le Journal de la Scribouilleuse: Un pseudonyme qui cache un auteur qui nous narre au jour le jour l’ordinaire et l’extraordinaire d’une jeune femme parisienne. Le grand défouloir d’une scribouilleuse tantôt profonde tantôt délicieusement superficielle !” (Links page of a personal site)

The home page of the Journal de la Scribouilleuse (“Le journal de bord de la Scribouilleuse”) contains, among other items, the following “dishes” (the various sections are presented in the form of a menu):

“La Scrib au Québec [...] Le composite de la Scribouille
[...] La revue de presse de la Scrib”

Other words to be added to the list of members of the lexical family of *scribouiller*: a) **scrib**, n. f. Abbreviation of **scribouilleuse** ; b) **scribouille**, n. f. Nickname used in interactive narratives:

“La Scribouille m’a avoué que malgré qu’il lise parfaite-

ment le latin, souvent le contenu des textes qu'il recopiait pour nos maîtres restait incompréhensible. Et Dame Forgaill m'a expliqué qu'avant de les comprendre il fallait faire de très longues études avec des gens comme Maître Hérasmus ou Maître Radamanth. Je pense que je préfère avoir la Scribouille comme professeur."

As for **scribouillage** (67 occurrences found in June 2003), the majority of occurrences of the singular have the pejorative connotation of the dictionary:

"[J]'espère que vous mesurez la chance que vous avez eu d'échapper à cet infâme scribouillage." (*Le Blog de Torpedo*)

including an echo from the past:

"L'important est que la mère et l'enfant soient réunis et que l'autre fillette puisse rejoindre sa famille dans les meilleurs délais, le reste n'est que grenouillage et scribouillage." (*Observatoire de l'Afrique centrale*)

The plural indicates clearly that it is a question of the result of the action of *scribouiller* rather than the action itself:

"*Barbouillages et Scribouillages*", title of a scenario (*La Scénariothèque*)

"Voilà mes petits scribouillages, bonne lecture à vous, mesdames." (Page of *Scribouillages*)

The word family of *scribouiller* shows then two faces on the Web: on the one hand, in the background, the reflection of the dictionary *scribouiller*, and on the other, in the foreground, a form which, by its onomatopoeic expressivity portraying a contemporary content, suits admirably well the "underground" world -- non-academic and especially non-"literary" -- of online interactive writing: the world of *scribouilleuses*, *scribouillards*, *scribouilleux*, *scribouilles* and *scribs* who *scribouillent* their *scribouillages* in *scribouillons*, *scribouillards* or *scribouill'arts*.

3. Syntagmas and paradigms

3.1. *Esprit de corps, esprit d'équipe*

These expressions are semantically closely related (parasyonyms). The Petit Robert privileges the former, implicitly suggesting that it is more lexicalised, or more frequent, than the latter:

“*Esprit de corps*, d’attachement et de dévouement au corps, au groupe auquel on appartient. => **corporatisme, solidarité**. *Esprit d’équipe*.” (s.v. ESPRIT)

The *Dictionnaire des expressions et locutions figurées* (Robert, 1979) gives more details:

“***Esprit de corps***, «sentiment de solidarité entre les membres d’un groupe» (depuis 1767, in Wartburg). ***Esprit d’équipe*** a les mêmes emplois dans un contexte plus précis [...]” (s.v. ESPRIT)

Esprit d’équipe is not dated, but the dictionary suggests implicitly that it is more recent than *esprit de corps*. The Web shows clearly the predominance of *esprit d’équipe* in a present-day world demanding more competitiveness (*esprit d’équipe* imposed from above) than solidarity (*esprit de corps* between equals):

Google, French-language pages (December 2002): *esprit de corps* x 4,130 ; *esprit d’équipe* x 28,800.

One need look no further that the first results offered by Google to see that *esprit d’équipe* is a key term/concept:

Page titles: “**Avez-vous l’esprit d’équipe?**” ; “**Un pour tous, tous pour un. Favoriser l’esprit d’équipe.**” ; “**Esprit d’équipe.**”

Web address (URL): <<http://www.esprit-equipe.com>> ; including the meta tag “<META NAME=’Description’ CONTENT=’L’Esprit d’Equipe est une agence de

marketing sportif organisée autour de 4 métiers complémentaires: les Relations publiques, le Sponsoring, l'Événementiel et la Promotion de marques.”>”

3.2. *Madame la ministre / madame le ministre*

A few years ago it was possible to note from a media corpus that contemporary French accepts more and more titles such as *Madame la ministre*. Without the need to undertake the preparatory work of collecting a corpus of newspaper articles, one can observe at a glance the following distribution in French-language documents on the Web (end of September 2003):

1. Google, Web: *la ministre* x 400,000
 - a) Google, Web: *madame la ministre* x 21,900 + *mme la ministre* x 14,300 = 36,200 = 77.00 % of a+b
 - b) Google, Web: *madame le ministre* x 5,280 + *mme le ministre* x 5,530 = 10,810 = 23.00 % of a+b

2. Google, Canadian pages (.ca): *la ministre* x 121,000
 - a) Google, Canadian pages (.ca): *madame la ministre* x 2,660 + *mme la ministre* x 2,220 = 4,880 = 98.87 % of a+b
 - b) Google, Canadian pages (.ca): *madame le ministre* x 38 + *mme le ministre* x 18 = 56 = 1.13 % of a+b

3. Google, French pages (.fr): *la ministre* x 137,000
 - a) Google, French pages (.fr): *madame la ministre* x 11,300 + *mme la ministre* x 8,250 = 19,550 = 72.70 % of a+b
 - b) Google, French pages (.fr): *madame le ministre* x 3,090 + *mme le ministre* x 4,250 = 7,340 = 27.30 % of a+b

4. Google, Belgian pages (.be): *la ministre* x 12,100
 - a) Google, Belgian pages (.be): *madame la ministre* x 1,230 + *mme la ministre* x 359 = 1,589 = 93.97 % of a+b
 - b) Google, Belgian pages (.be): *madame le ministre* x 80 + *mme le ministre* x 22 = 102 = 6.03 % of a+b

5. Google, Swiss pages (.ch): *la ministre* x 2,000

- a) Google, Swiss pages (.ch): *madame la ministre* x 123 +
mme la ministre x 30 = 153 = 80.53 % of a+b
- b) Google, Swiss pages (.ch): *madame le ministre* x 28 +
mme le ministre x 9 = 37 = 19.47 % of a+b

These raw figures indicate: on the one hand, that the feminine (at least in the context *Mme/Madame la ministre*) is the more frequent in French-speaking countries in general and in four francophone countries in particular ; and on the other hand, that it has an above-average frequency in Canada, Belgium and Switzerland (almost exclusive in Canada), but below average in France.

3.3. *Se moquer de qch/qn comme de sa première chemise*

This sort of expression allows variable formulation (*Se moquer de X comme de sa/son première/premier Y*): a verb belonging to the semantic field of *se moquer de*, a noun belonging to that of the first object possessed by the subject of the verb. The queries “**comme de * premier**” and “**comme de * première**” produce, among others and with little noise (Google shows them in its display of results -- September and December 2002):

Verbs: *m'en tape, faire fi, m'en fous, s'en balancent, m'en tamponne, s'en carre, se fiche de, s'en crissent, se soucie de, s'en occuper, s'en préoccupent, s'en contrefout, s'en bat l'oeil.*

The nouns vary even more, often being chosen in relation to the theme of the context ; for example:

- *bulle* in speaking about a papal injunction (“les Américains se moquent, comme de leur première bulle, de cette injonction papale”),
 - *perruque* in speaking about Vivaldi (“Vivaldi, évidemment, s'en fiche comme de sa première perruque”),
 - *moustache* in speaking about Georges Brassens (“dont Brassens se foutait comme de sa première moustache”).
- One notes however that the normal (expected) noun

chemise is used 18 times ; *chaussette* occurs four times.
Among other more or less “normal” first objects: *paire de culotte* (or simply *culotte*, *culotte courte*, *couche-culotte*, or even *Sainte Culotte*), *paire de bobettes*, *sucette*, *socquette*, *robe*, *slip*, *caleçon*, *anorak*, *cotillon*, *biberon*, *cure-dents*, *kleenex*, *cheveu*, *baiser*.

The results obtained on the Web show the high degree of productivity of the expression.

4. Homophonic, paronymic, morphological confusions

This section is not without interest to the French teacher trying to understand learners’ mistakes.

4.1. *Sous les meilleurs auspices / sous les meilleurs hospices*

The confusion is so common that it is not surprising the the Web site of the television station France 2 put up the second version (until a reader pointed out the mistake!).

Google, French-language pages (November 2002): a) *sous les meilleurs auspices* x 2,970 = 88.45 % of a+b ; b) *sous les meilleurs hospices* x 388 = 11.55 % of a+b.

Homophonic confusion is not limited to the last word of the expression:

Google, French-language pages (November 2002): *sous les meilleures auspices* x 268 ; *sous les meilleures hospices* x 96.

Not to mention *hauspices* or *ospices*.

4.2. *Conjecture/conjoncture*

The paronymic mistake dans *la conjecture actuelle*, for *dans la conjoncture actuelle*, appears to be far less frequent than the homophonic mistake just mentioned:

Google, French-language pages (April 2003): *dans la conjoncture actuelle* x 2,590 ; *dans la conjoncture actuelle* x 15.

On the other hand the confusion *se perdre en conjectures / conjonctures* is less uncommon:

Google, French-language pages (April 2003): *se perd/perdent/perdre en conjectures* x 1,030 ; *se perd/perdent/perdre en conjonctures* x 72
AltaVista, French-language documents (April 2003): *se perd* en conjectures* x 214 ; *se perd* en conjonctures* x 22

The occurrences of the incorrect form correspond almost exclusively to a real confusion ; there is only one exception in Google:

“On ne dit pas... ‘La police se perd en conjonctures.’
On doit dire... ‘CONJECTURES’ *Hypothèses..*” (Page Apprenez à bien parler le français)

4.3. *Plurals in -x*

This is a complex zone, often grey, of French grammar. There are first of all accepted variants:

cf. *TLFI*: “En fait, il y a flottement entre *finals* et *finaux*: le 1er semble être le plur. de la lang. cour. et des écrivains, le second celui des linguistes et des économistes; ex.: *des b., d, g finaux, les résultats finaux* (cf. DUPRÉ 1972, pp. 1011-1012).”

Google’s French-language pages (April 2003) show the following figures:

résultats finals x 1,250 ; *résultats finaux* x 4,360
rapports finals x 1,120 ; *rapports finaux* x 1,800
examens finals x 719 ; *examens finaux* x 1,960

chantiers navals x 19,400 ; *chantiers navaux* x 57
combats navals x 685 ; *combats navaux* x 10

prénatals x 3,980 ; *prénataux* x 2,950
postnatals x 1,500 ; *postnataux* x 685

Then there are the well-known -oux:

bijous x 487 ; *bijoux* x 226,000
caillous x 645 ; *cailloux* x 52,100
chous x 1,170 ; *choux* x 77,200
genous x 1,620 ; *genoux* x 396,000
hibous x 345 ; *hiboux* x 10,900
joujous x 486 ; *joujoux* x 4,960
pous x 16,400 ; *poux* x 27,200

These figures, like all raw figures, need to be refined, of course. For example, the surprising figure of 16,400 occurrences of *pous* includes a certain number of proper nouns and typos (*pous* instead of *pour*). We shall add the occurrences of the plural of the adjective *chou*:

cf. *Petit Robert*: “**Adj. inv.** *ce qu’ils sont chou.*”

On the Web (Google, French-language pages), we found in April 2003:

ils sont chou x 24 ; *ils sont chous* x 16 ; *ils sont choux* x 45

To finish with, we shall look at what would be considered mistakes by normative grammar:

“Parc d’activités des Travaux, Cagnes sur Mer” (*Sites d’activités des Alpes Maritimes*)

“PROGRAMMES NATIONALS DE FORMATION DU PROJET DISNET-STEP” (Page of the Web site of l’Institut Universitaire de Technologie, Université de Bordeaux I)

5. Canadian and French usage

5.1. *Courriel* vs. *email*

The semantic field of *courriel/email* is fairly difficult to cover statistically. We shall approach it from two Web sites, one Canadian, the other French:

“Anglais *electronic mail* Synonyme(s): *e-mail*. Français *courrier électronique* n. m. recommandé par l’Office de la langue française [...] Synonyme(s): *courriel* n. m., *messagerie électronique* n. f.” (Vocabulaire d’Internet de l’Office québécois de la langue française, September 2003)

“Anglais *electronic mail message* Synonyme(s): *e-mail message, electronic message, e-mail, e-message* [...] Français *courriel* n. m. [...] Synonyme(s): *courrier électronique* n. m., *message électronique* n. m. Terme(s) à éviter: *e-mail*” (Ibid.)

There is one distinction and several confusions. The Office makes a distinction between the virtual system -- for which it recommends *courrier électronique* and gives two synonyms, including *courriel* -- and its punctual realization -- for which it gives, without making a recommendation, *courriel*, the synonyms *courrier électronique* and *message électronique* and the proscription of *e-mail*. Polysemy then of *courriel* and *courrier électronique*. In real usage, it is the meanings of “message électronique” and “adresse électronique” that predominate and invoke the simple forms, as can be seen on an address page of l’Ecole Doctorale Chimie et Sciences du Vivant de l’Université Joseph Fourier de Grenoble which contains six distinct forms:

e-mail, email, mail, mel, mél and *mèl*

These two sites -- the prescriptive Office (*langue*) and practical Grenoble (*discours*) -- would seem to suggest that the simple form *courriel* is Canadian, whereas the variants *e-mail, email, mail, mel, mél* and *mèl* are French. How can one confirm this impression on the Web ? We shall put aside the form *mel*, too often a proper noun (*Mel Gibson, Mel’cuk*, etc.), and take

into account the fact that for Google the query **e-mail** corresponds indifferently to *e-mail* and *email*, that **mél** corresponds indifferently to *mél*, *mel* or *mèl*. Our observation is limited to explicitly Canadian (**.ca**) addresses on the one hand, and French (**.fr**) on the other, in the French-language pages of Google (April 2003).

Query **courriel site:.ca**: *courriel* x 248,000

Query **courriel site:.fr**: *courriel* x 56,000

“Query +”**e-mail**” ” -**email site:.fr**: *e-mail* x 1,000,000

“Query +”**e-mail**” ” -**email site:.ca**: *e-mail* x 81,000

“Query +”**e-mail**” ” -”**e-mail**” **site:.fr**: *email* x 624,000

“Query +”**e-mail**” ” -”**e-mail**” **site:.ca**: *email* x 54,400

Query +**mél -mel site:.fr**: *mél* x 103,000

Query +**mél -mel site:.ca**: *mél* x 811

Query +**mèl -mel site:.fr**: *mèl* x 6,870

Query +**mèl -mel site:.ca**: *mèl* x 29

In short, *courriel* is much more frequent on Canadian sites, *e-mail*, *email*, *mél* and *mèl* on French ones. The forms *e-mail* and *email* should doubtless be looked at in more detail since they are often encountered in English sequences contained within what Google considers French-language documents.

5.2. *Pourriel* vs. *spam*

After *courriel*, *pourriel*. The latter word, which l’Office québécois de la langue française insists on distinguishing (pointlessly, in our opinion) from *pollurriel* (*pourriel* = Eng. *junk e-mail* ; *pollurriel* = Eng. *spam*), is, like *courriel*, of Canadian origin. Among the many references in online Quebec media, we shall quote André Forgues:

“**Le monstre du pourriel** (André Forgues). Selon des statistiques récemment publiées, chaque internaute recevait 571 messages de *pourriel* en 2001 et il en recevait 1500 en 2006. Cette croissance du nombre de messages de

courrier électronique non sollicités et non désirés explique en bonne partie d'autres prévisions selon lesquelles le volume total de messages transmis par Internet va passer de 31 milliards cette année à 60 milliards en 2006 ! [...] Et il y a les chefs d'entreprise qui commencent à se soucier des pertes de productivité engendrées dans leurs organisations par cette avalanche de pourriel, ainsi que les "honorables" sociétés de commercialisation qui s'inquiètent de voir un formidable outil de communication risquer de leur échapper, les internautes faisant de moins en moins la différence entre les envois commerciaux "légitimes" et le pourriel, ou spam." (*Le Soleil*, 7 October 2002)

Forgues had talked about *pourriel* in 1999 in *Cyberpresse* ; l'Office québécois de la langue française had already made proposals for the naming of the phenomenon in 1997. The *Jargon français* notes the origin of the term in 1999:

"**pourriel** n. m. [ARGOT] Néologisme, que j'ai trouvé sur quelques sites québécois, formé par la contraction de « courriel pourri », et désignant bien évidemment le spam." (*Le Jargon Français*, 27 February 1999)

What does the Web say in April 2003 ?

Google, French-language pages: *pourriel* x 2 100, *pourriels* x 12 000

Google, French-language pages: *spam* x 127 000

Google, French-language pages, Canadian sites (.ca): *pourriel* x 341, *pourriels* x 251

Google, French-language pages, French sites (.fr): *pourriel* x 123, *pourriels* x 69

Google, French-language pages, Canadian sites (.ca): *spam* x 3,990

Google, French-language pages, French sites (.fr): *spam* x 18,500

In short: a) the plural *pourriels* is more frequent than the singular *pour-*

riel in the global French-language Web -- it concerns concrete objects --, whereas on national servers the singular, the general phenomenon, predominates ; b) *pourriel(s)* is more frequent on the Canadian Web than on the French Web ; c) the anglicism *spam* is by far the most frequent word in all cases.

6. Genesis of the concept “une autre mondialisation” and its lexical denomination by *altermondialisation*

With the Larzac gathering preceding the WTO conference in Cancun, the French press of August 2003 made a higher than normal use of the terms *altermondialisation*, *altermondialiste(s)* and *altermondialisme*. For example:

“La rentrée seule dira si le cadre festif et symbolique du Larzac 2003 aura été autre chose qu’un Woodstock altermondialiste. La cohérence n’est pas évidente entre la lutte des intermittents et celle pour la fin des subventions agricoles. Il n’est pas certain que le credo altermondialiste puisse servir de tronc commun aux contestations hétérogènes et multiformes qui ont fusionné un chaud week-end d’août sur la planète Larzac.” (*Libération*, 11 August)

“ALTERMONDIALISATION: Philippe Seguin estime que les altermondialistes ne proposent que la suppression de l’OMC.” (*Le Monde*, 19 August)

Altermondialisation is the phenomenon ; *altermondialisme* is the ideology ; the *altermondialistes* are the followers of the *altermondialiste* movement. In September 2003, Google’s French-language pages gave the following statistics:

altermondialisation x 6,730 + *alter-mondialisation* x 1,740 = 8,470
altermondialisme x 1,770 + *alter-mondialisme* x 202 = 1,972
altermondialiste(s) x 19,260 (8,660 + 10,600) + *altermondialiste(s)* x 3,470 (1,180 + 2,290) = 22,730

To what extent does the Web allow for the dating of the concept and the lexical term ? We shall first of all quote from an article of the Wikipédia:

“Altermondialisation

Un article de Wikipédia, l’encyclopédie libre.

L’alter-mondialisation est un terme qui désigne une forme de contestation politique apparue à la fin des années 1990.

Les **altermondialistes** se posent en opposants du néo-libéralisme, de la guerre, de la pollution, des fascismes, de la pauvreté, etc. L’alter-mondialisation regroupe des personnes d’horizons divers, des nationalistes qui regrettent le délitement de l’idée d’État-nation, des démocrates qui constatent les limites de la démocratie, des collectivistes laissés orphelins par la chute du communisme, etc. Pour cette raison, cette nouvelle mouvance est souvent appelée *le mouvement des mouvements*.

Les manifestations de Seattle en 1999 contre l’OMC ont en quelque sorte lancé le mouvement des manifestations massives alter-mondialistes. Le rassemblement de Gênes en 2001, contre le sommet du G8 a vu la mort d’un manifestant tué par balle par la police italienne. En novembre 2002, s’est tenu le 1er Forum Social Européen à Florence, et 1 million de personnes ont défilé contre la guerre en Irak et *pour un autre monde*.” (“Altermondialisation”, Wikipédia, 22 August 2003)

The idea of a different world, of a different form of globalization would seem to date back to 1999. For the years 1999 and 2000, several attestations can be found, including:

1999 **“L’autre mondialisation.** Depuis quelques années, des voix s’élèvent pour dénoncer la « pensée unique » néo-libérale. Comment cette contestation est-elle devenue un phénomène planétaire ? Quelles sont les solutions alternatives ? Documentaire de François Christophe (France, 1999-1h). Enquête: Frédéric Castaignède. Coproduction: La Sept ARTE, Les Films du Village” (*Le Monde diplomatique*, “Transversales science-culture / La

mondialisation sur ARTE”)

9 May 2000 “09/05/2000: Premier café-citoyen organisé par ATTAC 66: « une alternative citoyenne à la mondialisation » (café la Source, Perpignan).” (Attac66)

12 June 2000 “**Congrès de la CISL: l’autre mondialisation** [...] Date de mise à jour du document: 12/06/00” (Union Interprofessionnelle de la Réunion)

In September 2003, the following frequencies could be observed for certain key expressions (French-language pages, Google):

pour une autre mondialisation x 4,990
une autre mondialisation est possible x 610
une autre mondialisation est nécessaire x 15
une autre mondialisation en mouvement x 175
une autre mondialisation en marche x 14

We shall now look at online attestations of *altermondialisation*, *altermondialiste* and *altermondialisme*.

altermondialisation

3 January 2002 “**Manifeste Initiatique Révolutionnaire** 3 janvier 2002 [...] Dans le mouvement politique d’altermondialisation, l’utilisation de sociétés secrètes, de la propédeutique & du modus operandi de l’occultisme peuvent devenir des outils de lutte.” (Manifeste Initiatique Révolutionnaire)

21 March 2002 “21 mars 2002. Un pays décomplexé (4/5) L’Hexagone montre qu’il est vital de résister ». DE FILIPPIS Vittorio; LOSSON Christian; MUSSEAU François. [...] de la capacité d’écoute de la France vis-à-vis des militants de l’altermondialisation et de sa ténacité dans la défense de ses intérêts.» (Earliest occurrence in Archives of *Libération*)

April 2002 “**SOMMAIRE du n°73, avril 2002.**
Page 10 Actuplus - Porto Alegre ou les formes de
‘l’altermondialisation’” (*Science Frontières*)

8 June 2002 “Edition du 8 juin 2002. Jospin, retour
sur une déroute [...] « *La social-démocratie ne fait
qu’accompagner la mondialisation en donnant le sen-
timent qu’elle peut de moins en moins agir sur elle* »,
objecte Yves Salesse, président de la Fondation Copernic
et partisan de l’« altermondialisation » (terme préféré à
celui d’antimondialisation, considéré comme trop négati-
f).” (Earliest occurrence in Archives of *Le Monde*)

altermondialiste, adj.

2 February 2002 “**L’histoire de l’irruption du mouve-
ment altermondialiste.** [...] Article paru dans l’édition du
2 février 2002.” (*L’Humanité*)

13 September 2002 “**La société civile et le politique.**
Quels dialogues, quels rapports entre la société civile et
le pouvoir politique? Quelles convergences, quelles com-
plémentarités entre mouvements sociaux et politiques «
altermondialistes»? Quelle articulation entre démocra-
tie représentative et démocratie directe? 13/09/2002 -
Horman Denis - Article” (Attac France)

21 November 2002 “**La lente marche du réformiste
Hollande** [...] Alors que les partisans du Nouveau Monde
lorgnent vers les mouvements « altermondialistes », Hol-
lande a revendiqué « une confrontation avec l’extrême
gauche” (*Libération*, 21 November 2002)

altermondialiste, n.

18 December 2001 “**Les Legal Teams: dernier rempart
de sécurité des altermondialistes.** Olivier Coussaert, 18
décembre 2001. Anti-mondialiste, Alter-mondialiste,...
Très bon travail des Legal Team! Mais ma remarque porte

sur le titre: voila donc qu'entrent en scène les altermondialistes. Je suppose que c'est une tentative de « positiver » le mouvement. Anti, cela fait négatif, n'est-ce pas?" (Page of forum entitled *Votre opinion...*)

21 March 2002 “**Un pays décomplexé (4/5) L'Hexagone montre qu'il est vital de résister** [...] l'image corporatiste, ringarde et très franco-française de la manifestation traditionnelle et institutionnelle, les altermondialistes partisans d'une autre mondialisation made in France figurent parmi les pionniers” (*Libération*, 21 March 2002)

17 October 2002 “Lettre ouverte aux alter-mondialistes. « L'hypocrisie derrière notre compassion ». 17 octobre 2002 [...] Guy Verhofstadt” (On the site *Belgium at the United Nations*)

altermondialisme

23 July 2002 “**Altermondialisme ou reconstruction de la gauche ?** PAR BERNARD GUIBERT, mardi 23 juillet 2002” (AlphaVert.net)

18 March 2003 “18 mars 2003. **Economie** [...] modèle de cybermilitantisme tous azimuts. Entre contributions et appels, décryptages et plaidoyers, Babel de l'altermondialisme.” (Earliest occurrence in Archives of *Libération*)

1 June 2003 “LES MILLE VISAGES DE L'ALTERMONDIALISME. Article publié le 1 Juin 2003. [...] Ils refusent la « gouvernance mondiale » des pays les plus riches, dénoncent les inégalités et réclament d'autres modèles. Par dizaines de milliers, les altermondialistes se sont donné rendez-vous dimanche autour du lac Léman pour faire entendre, dans toute leur diversité, leurs voix face aux dirigeants du G8.” (Earliest occurrence in Archives of *Le Monde*)

These results, obtained in a few minutes of querying, show that the first attestations of this word family can easily be dated back to the beginning of the year 2002, or even earlier. According to these first data, the word *altermondialisme* seems to have appeared after the other two, which is logical: *altermondialisation*, having replaced *antimondialisation* in order to present a more positive opposition to *mondialisation*, later became an ideology, *l'altermondialisme*.

7. Dictionary selectivity and Web extensivity: *doudou*

To illustrate this point (we have already done so in passing in several of the preceding sections), we shall take as our starting point Richard Desjardins' song *Caroline*:

“Prends ta robe et ton bijou, / dis bye bye à ta doudou”

What then is a *doudou*? We first of all look in a few dictionaries:

Petit Robert, 1993

DOUDOU [...] n. f. [...] RÉGION. (Antilles) Jeune femme aimée. *Il était avec sa doudou.*

Trésor de la langue française informatisé

DOUDOU, subst. fém. *Région.* (Antilles) et *fam.* Appellation tendre donnée à une femme. *L'amour (...) La prononciation la trahit (...) la syllabe mour, prononcée comme mou, presque sans r et avec une ferveur de doudou* (H. BAZIN, *Lève-toi*, 1952, p. 111).

Dictionnaire universel francophone en ligne

doudou n. (Antilles fr.) Fam. Chéri(e).

Encyclopédie Voilà avec Hachette

doudou n. fém. *Fam.* Femme ou mère,
aux Antilles.

If there is a certain amount of confusion, we can say that globally *doudou*, a feminine noun, is a colloquial, affectionate Caribbean term designating a woman. Is Desjardins, a Quebecer, then telling the young Caroline to leave her dear (Caribbean) woman ? It hardly seems likely.

Let us then look elsewhere, that is to say on the Web. The Swiss singer Henri Dès says on the subject of his song *Mon doudou*:

“Dans mon « doudou » je parle d’un chiffon que les enfants ont à une certaine époque de leur vie et je sais qu’il y a beaucoup d’adolescents qui le mettent sous leur lit parce qu’ils ne veulent pas le perdre, même s’ils n’osent dire qu’ils en ont encore un... pour ne pas paraître ridicules aux yeux de leurs copains.” (Page entitled “Une âme de gamin”, on the site of the newspaper *L’Express à Neuchâtel*)

We can now propose the hypothesis that one says *une doudou* in Canada, *un doudou* in Europe.

To turn to raw statistics (Google, French-language pages, May 2003):

doudou(s) x 29,270
mon/ton/son doudou x 2,490
ma/ta/sa doudou x 733

For the feminine, there are occurrences of both meanings, that of the dictionaries and that of Desjardins and Dès. The second is the one which interests us at present. We shall quote two pages:

“Tu sais, même les grands sont jaloux / de toi ma Doudou
et par-dessus tout, / il va bien falloir / que tu te décides
à apprendre à nager, / car maman veut te rapiécer / et te
plonger dans la machine à laver.
Le soir ce que j’aime le plus, / c’est de m’endormir auprès
de toi. / J’aime ta senteur parfumée de fleurs, / quand tu es
tout enroulé entre mes doigts / et que tu me chatouilles le

visage. / Tu es mon ami, tu me réconfortes beaucoup.
Je t'aime plus que tout, toi ma Doudou.” (Extract from
“Regards des petites personnes: Ma doudou”, Canadian
page)

“Ce ne sont pas les parents qui choisissent le doudou mais
les enfants. [...] Peluche, morceau de tissu ou objet, le
doudou les rassure.” (Extract from “Le premier doudou”,
French page)

Un or *une doudou* is then a stuffed animal or a piece of cloth chosen by the infant as the inseparable friend who brings comfort. Children's language -- a large part of which disappears when the infant reaches school age, to return with parenthood -- is far more present on the Web than in dictionaries.

Conclusion

Macro-lexicography is today based on the analysis of costly and essentially static large corpora, thus obsolescent in respect of actual usage, that of the moment of consultation of the dictionary. The observation of the Web as a corpus of linguistic usages, situated as it is at the level of micro-lexicography (traditionally the domain of lexicology or word studies), has the advantage of being able to renew itself constantly by being based on dynamic corpora.

More detailed analyses of the phenomena presented in this paper can be found in the Net des Études françaises <<http://www.etudes-francaises.net/acre/corpus/>> or the mirror site <<http://translatio.ens.fr/miroir-nef/acre/corpus/>>.

Works Cited (dictionaries only)

- Dictionnaire des expressions et locutions figurées*, Paris: Robert, 1979.
Dictionnaire québécois d'aujourd'hui, Saint-Laurent: DicoRobert, 1993.
Petit Robert, Paris: Dictionnaires Le Robert, 1993.
Trésor de la langue française informatisé, <<http://zeus.atilf.fr/tlf.htm>>.