

Gilles SIOUFFI, Université Paris-Sorbonne  
Agnès STEUCKARDT, Université Paul-Valéry Montpellier III  
Chantal WIONET, Université d'Avignon

## Comment enquêter sur des diachronies courtes et contemporaines ?

### 1 Introduction

Dans son histoire, l'un des principaux problèmes qui se sont posés à la linguistique historique – plus spécifiquement diachronique – est que l'exploitation manuelle des corpus écrits rendait plus facile l'observation et l'analyse de phénomènes récurrents au sein de ce qu'on pourrait appeler des « synchronies historiques » (un siècle, par exemple, ou l'empan de l'« ancien français ») que de phénomènes touchant proprement au changement linguistique. Les études diachroniques se sont ainsi construites dans la mise en évidence de grandes évolutions, plus que dans le repérage des moments particuliers où quelque chose dans la langue *change*.

Aujourd'hui, deux phénomènes liés à la recherche sont venus singulièrement modifier la donne. Le premier est l'interrogation automatique des corpus, laquelle permet de mettre au jour, à partir de la sollicitation d'une très grande quantité de sources autour d'un même objet – un mot, par exemple – des faits d'émergence et de diffusion qui étaient difficiles à repérer lorsque l'on s'appuyait sur le seul dépouillement en continu de sources écrites lues. Le second, lié à la vogue de la problématique du changement linguistique, est que l'on ne privilégie plus nécessairement les diachronies longues dans l'enquête. Ainsi, la diachronie courte a-t-elle acquis une nouvelle légitimité, dans l'étude du lexique de spécialité (Picton et Condamines, 2010), d'abord, mais aussi du lexique général. En effet, c'est bien sur des diachronies courtes que les phénomènes d'innovation peuvent être cernés. Pour autant, ce type de recherche ne peut être mené que si on dispose d'une quantité suffisante de données, et, qui plus est, bien contextualisées.

Enfin, on observe que, si les enquêtes diachroniques ont jusqu'à présent surtout porté sur des périodes anciennes, l'approche diachronique d'un français récent, voire très récent, n'est plus considérée comme illégitime (Combettes, 2011). Ces états de langue ne sont plus nécessairement compris comme construisant une « synchronie » du français moderne (Narjoux éd., 2011), comme cela a pu être le cas quand, peut-être sous l'influence de l'objectif pédagogique, l'étude diachronique était avant tout mobilisée lorsqu'il y avait rupture de la compréhension. Certes, enquêter sur le français contemporain sous l'angle diachronique peut avoir en soi quelque chose de paradoxal, mais prend son sens précisément si l'on cherche à approcher de plus près les dynamiques entourant l'émergence d'un fait nouveau.

Cette manière de comprendre la diachronie et la spécificité du terrain d'étude engagent-elles dès lors de nouvelles méthodologies ? La présente communication se concentrera sur quelques interrogations et difficultés qui peuvent apparaître lorsqu'on mène une recherche autour d'usages émergents en français contemporain, et lorsqu'on cherche à les saisir en diachronie. Elle soulèvera trois types de questions qui peuvent naître au moment d'initier semblables enquêtes :

- Comment isole-t-on le fait sur lequel va porter l'investigation ?
- Quels types de corpus interroger ? Quel corpus est susceptible de donner les meilleurs résultats ? Comment interroger ?
- Que signifie « étudier un usage émergent en diachronie » ? Doit-on se contenter du repérage de l'innovation ? Doit-on / peut-on dépasser ce stade ? Comment enquêter sur la diffusion ?

Les réflexions qu'on lira ici, de nature essentiellement méthodologique et programmatique, s'insèrent dans le cadre d'une enquête menée sur des objets spécifiques : les modalisateurs en français contemporain, parmi lesquels notamment le confirmatif *c'est vrai que* (type *c'est vrai que c'est bien*) et l'intensif *juste* (type *juste incroyable, juste magnifique*) qui serviront ici à illustrer le propos, et qui ont fait l'objet d'une investigation sur une diachronie très courte et très récente (2000-2010) et sur des corpus écrits (la question de la prise en compte de l'oral ne sera pas du tout évoquée ici). L'objectif est de comprendre si, à partir des attestations écrites que l'on aura recueillies, les faits en question relèvent de simples « modes langagières » (voir Siouffi, dir, à paraître), autrement dit si leur étude en micro-diachronie fait apparaître à peu de temps de leur émergence des signes de retrait, ou s'ils sont les prémices d'un véritable *changement*, durable dans la langue.

## 2 L'« événement méta »

On le sait, si certains faits langagiers nouveaux se caractérisent d'emblée par une forte visibilité (néologie terminologique, emprunt...), d'autres, particulièrement dans les domaines phonétique, sémantique et syntaxique, sont parfois plus difficiles à repérer. Il faut, comme on dit, tendre une oreille fine pour percevoir des micro-différences dans l'ouverture d'un *a*, dans les nuances d'emploi d'un mot de haute fréquence, dans la sous- ou sur-représentation d'une construction. Dans la forêt des usages linguistiques, sur quoi donc s'appuyer pour distinguer ceux qui peuvent être qualifiés d'*émergents* (et ici, nous ne ferons pas de distinction, pour le moment, entre l'*émergence* et l'*innovation*) ?

Rêvons une enquête statistique idéale : une fois constitués deux échantillons représentatifs de la langue commune, l'un pour l'instant  $t_0$  l'autre pour l'instant  $t_1$ , on mesurerait les écarts de fréquence, au moins entre formes lexicales, et éventuellement, par le biais des annotations, entre d'autres caractéristiques linguistiques. On pourrait ainsi quantifier les augmentations récentes et significatives de fréquence entre l'état actuel la langue commune et l'état immédiatement antérieur. Un tel projet reposerait malheureusement sur des présupposés contestables : que pourrait être en effet « un échantillon représentatif de la langue commune » ? Autant de corpus, autant de variétés des usages. Les corpus, si grands soient-ils, ne donnent de la « langue commune » qu'une représentation biaisée, qu'ils reflètent l'usage littéraire, comme Frantext, l'usage journalistique, comme les bases de données de presse, ou encore des variétés particulières de l'usage oral, comme par exemple les corpus oraux réunis dans la banque de données « Parole et langage » (<http://sldr.org>). Même dans l'espace du web, dont l'étendue semble repousser la limitation inhérente aux traditionnels corpus, il n'est pas assuré que l'abondance et la diversité suffisent à annuler les effets de surreprésentation, à équilibrer les disparités, puisque l'expression sur ce médium induit par elle-même des choix stylistiques particuliers. Une manière de surmonter cette difficulté est sans doute d'améliorer la représentativité des corpus, en constituant d'abord de grands corpus, comme il a été entrepris au sein de l'infrastructure de recherche « Corpus », avec les « Consortium Corpus Oraux et Multimodaux » et « Consortium Corpus écrits », pilotés par le CNRS.

Cependant, la langue commune n'est pas un objet observable du même type que la population ou le climat. Plutôt qu'un ensemble de données tout à fait objectives, elle apparaît comme un objet virtuel, présent dans l'horizon conceptuel des locuteurs, mais qui ne peut être complètement réalisé par aucune somme de discours concrets. C'est du moins sur cette conception de la « langue » que nous nous appuyons pour initier la recherche que nous présentons ici. Parce que la langue commune est avant tout un objet d'usage, la prise en compte de la perception qu'en ont les locuteurs ordinaires pourra être vue comme un mode d'accès possible et légitime à cette caractérisation. Sans confondre le plan des usages avec celui des représentations, on pourra partir de l'idée que, parmi les usages potentiellement descriptibles comme émergents, une partie aura d'abord été identifiée par les locuteurs eux-mêmes. Bien sûr, la question de la nature des faits visés – phonétiques, lexicaux, sémantiques, syntaxiques, pragmatiques- importe ici. Le fait que notre enquête porte sur des modalisateurs du dire a sans doute ses implications. Toujours est-il que, dans ce cadre précis et ces précautions exprimées, nous avons opté pour que ce soit d'abord le signalement par les locuteurs d'un usage qu'ils perçoivent comme émergent qui

serve de point de départ. Les corpus disponibles permettent de confronter, *a posteriori*, les représentations linguistiques des locuteurs ordinaires à des usages attestés ; ils ont, dans notre enquête, statut d'échantillons-témoins (voir partie 2).

Pour le français contemporain, où recueillir ces signalements ? Bien sûr, la conversation privée, en particulier avec des linguistes vigilants, peut en fournir de remarquables. Cependant le signalement privé, par définition, n'a pas une grande diffusion, et n'est donc pas susceptible d'avoir un impact sur l'attitude de la grande masse des locuteurs à l'égard de l'usage émergent, ni par conséquent sur l'évolution de sa fréquence. En revanche, pour peu qu'il circule dans les médias francophones, que ce soit la presse traditionnelle – notamment dans le courrier des lecteurs –, la radio, ou désormais le web, il appartient à ce domaine particulier de la vie sociale que constitue l'actualité linguistique. Si le journalisme grammatical ou les chroniques de langue semblent avoir beaucoup décliné depuis quelques décennies, encore que certaines individualités marquantes en prolongent la tradition<sup>1</sup>, l'existence du web fournit en effet à présent un nouvel espace d'expression au commentaire des innovations linguistiques.

Mais tous les signalements ne peuvent être placés sur le même plan : le commentaire isolé de quelque obscur blogueur n'a probablement pas, pour l'instant du moins, le même impact que l'article d'un chroniqueur de presse patenté. On peut ainsi hiérarchiser les signalements en fonction de leur fréquence et de leur diffusion ; et quand un signalement est remarquable par sa répétition ou par sa portée, on peut considérer qu'il fait événement dans l'actualité linguistique. Pour enquêter sur les usages émergents, il importe de repérer ce moment de ce que l'on nommera *l'événement méta*.

On attribuera ici à l'événement méta non seulement une valeur d'indicateur, mais aussi une valeur performative. En effet, quand bien même le locuteur ordinaire se fourvoierait, et signalerait comme émergent un usage déjà ancien ou moins fréquent qu'il ne le croit, le signalement importe en tant que tel : il met sur la place publique un fait de langue et, même si l'observateur se trompe, son observation fait partie de l'actualité linguistique. Nous partions ici de l'idée qu'un certain sentiment de la nouveauté a en soi son prix, en tant qu'il peut être, non seulement un facteur de modification des représentations, mais aussi des usages (voir Lecolle, à paraître).

Ce faisant, on réserve aux locuteurs ordinaires une place qui leur a parfois été refusée en linguistique. La perspective défendue ici n'est pas celle qui a longtemps prévalu par exemple en analyse de discours, où l'on a pu considérer que les outils d'analyse permettaient de détecter des faits linguistiques « invisibles à l'œil nu », inconscients du locuteur analysé, que l'on se plaisait, en somme, à « piéger ». On adopte ici, à l'égard des locuteurs, une position toute différente. À l'exemple des diachroniciens à l'écoute des « remarqueurs » (Caron éd., 2004), ou des lexicographes toujours preneurs de définitions naturelles (Ligas, 2008 : 42), on regardera les commentateurs ordinaires de la langue comme de vrais acteurs, comme des « sujets de langue », pour reprendre une expression de Sylvain Auroux (1986), et on estimera que leurs commentaires font partie des facteurs qui peuvent éventuellement influencer sur son devenir.

Cette attention qu'avec d'autres (Achard-Bayle et Paveau, 2008 ; Achard-Bayle et Lecolle, 2009) on prête ici aux commentateurs ordinaires n'implique évidemment pas adhésion à leurs commentaires. Ceux-ci, on a pu le constater à l'occasion de premières enquêtes menées sur les usages émergents en français dans la dernière décennie, sont éloignés d'une description sereine. Le signalement est souvent décliné sur le mode de la déploration, de la satire, ou de l'exaspération.

Ainsi, pour le modalisateur confirmatif *c'est vrai que* déjà émergent à la fin du 20<sup>e</sup> siècle, c'est dans la presse traditionnelle, sous la plume d'un chroniqueur de langue connu, que l'on a repéré le premier signalement. Il laisse paraître le mépris de ce remarqueur d'aujourd'hui :

On reconnaît [...] les rescapés de la bêtise ambiante, à ce qu'ils ne commencent pas toutes leurs phrases par *c'est vrai que* (Bertrand Poirot-Delpech, *Le Monde*, « On reconnaît », 24 juillet 1996)

Pour l'usage, plus récent, de *juste* comme adverbe intensif, c'est sur le web que l'on repère les premiers signalements. Dans l'interaction d'un forum, il est relevé en 2006, par un pseudo-nommé « Qui-Gon Jin », avec une note de fatalisme :

Le Pingouin : Mais LOL  
Qui-Gon Jin : Faut arrêter avec cette expression...  
Mon Colonel : Je suis d'accord ! ça vient d'où ce MAIS ?  
Qui-Gon Jin : Bah je sais pas... C'est un peu comme le « juste ».  
« C'est juste incroyable comment... » « C'est juste stupide blabla » Etc. Y a des expressions comme ça qui se propagent assez rapidement. (Forum FilmDeCulte, 31 octobre 2006, <http://forum.plan-sequence.com/les-films-qui-ont-coute-des-separations-des-ruptures-t4810-60.html#p81959>)

Sous les doigts d'une web-chroniqueuse, le commentaire sur *juste*, daté de la même année, est teinté d'ironie :

Quels ont été les mots, expressions ou tics de langage de 2006 ? Médaille d'or toutes catégories : "*C'est juste pas possible*", variante de "*j'veais juste pas avoir le temps*" ou plus positif, "*c'est juste hilarant*". Attention, toujours marquer un temps d'arrêt après "*juste*" : la fin du propos en sortira renforcée. (Sophie Lutrand, « Un target bankable, c'est juste pas possible », 26 décembre 2006, [http://lci.tfl.fr/france/societe/2006-12/target-bankable-est-juste-possible-4889224.html#voir\\_commentaires](http://lci.tfl.fr/france/societe/2006-12/target-bankable-est-juste-possible-4889224.html#voir_commentaires))

L'« événement méta » comporte ainsi deux aspects : il donne d'une part une information, dont la validité demande à être vérifiée, et, d'autre part, une appréciation, dont la portée mérite examen. Cette première étape de la recherche appelle donc deux contre-enquêtes, l'une sur les usages, l'autre sur la diffusion.

### 3 Méthodologie : le choix du corpus et son mode d'interrogation

Avec quel outil et sur quel terrain mener l'étude des usages émergents ? Lorsqu'il s'agit de formes anciennes, les corpus et bases de données traditionnels suffisent à dater leur apparition à l'écrit, à considérer leur fréquence, puis leur capacité à s'inscrire dans la langue. Mais pour une recherche menée en diachronie courte sur les nouveaux modalisateurs en français contemporain, le terrain est par définition à inventer. Ainsi par exemple, la base de données Frantext fournit seulement une dizaine d'occurrences du *c'est vrai que* confirmatif ; quant à l'usage intensif de *juste* devant adjectif, l'étude de son voisinage droit sur la période 2000-2010, ne fait apparaître aucun adjectif exprimant le haut degré. Dans une autre base de données, réputée « représentative », le *Corpus français* de l'Université de Leipzig, la séquence *juste + adjectif exprimant le haut degré*, épinglée par les nouveaux remarqueurs depuis 2006, n'apparaît pas<sup>2</sup> ; selon le logiciel d'exploitation associé à cette base de données, aucun adjectif ne fait partie des « voisins de droite significatifs » de *juste*. L'enquête doit donc s'attacher à élargir le terrain d'investigation.

D'autres bases de données du français écrit, interrogeables automatiquement, peuvent produire des résultats intéressants, comme *Europresse*, *Factiva*, *Lexis Nexis*, qui recueillent les articles des journaux écrits. Dans cette perspective, une enquête déjà menée sur *c'est vrai que* a permis d'élaborer une chronologie comparée des usages : on a pu mettre en évidence une augmentation de sa fréquence entre 2000 et 2004 dans la presse, des pics d'utilisation précoces (2004) dans *Libération* et dans *l'Équipe*, plus tardifs dans *Le Figaro* et dans *Ouest France* (2007), une décroissance générale de la fréquence à partir de 2007 (Siouffi, Steuckardt, Wionet, 2011). Globalement, si les pics d'utilisation de *c'est vrai que* n'apparaissent pas au même moment selon les journaux, la courbe que décrit *c'est vrai que* est nette : la forme est employée de manière croissante entre 2000 et 2004, pour connaître ensuite un déclin plus ou moins rapide selon les cas.

Pour l'unité lexicale complexe *c'est vrai que*, les bases de données de presse permettent de suivre au plus près, en diachronie courte, l'évolution de l'usage. Il n'en va pas de même pour un mot simple comme *juste* : la requête sur cette forme génère trop de bruit ; mais les collocations pointées par les remarqueurs permettent de démarrer l'enquête : ainsi la requête « *juste incroyable* », pointé en 2006 par le pseudo-nommé Qui-Gon Jin, ramène dans *Lexis Nexis* 110 occurrences, 168 dans *Factiva* et 239 dans

*Europresse*, avec une première attestation dès 1998<sup>3</sup>, une présence occasionnelle jusqu'en 2006 (une occurrence en 1998, une en 2001, deux en 2003), puis une montée en puissance de la fréquence :

| 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 |
|------|------|------|------|------|------|------|
| 1    | 4    | 14   | 41   | 34   | 66   | 61   |

Tableau 1. Nombre d'occurrences en valeur absolue de *juste incroyable*, entre 2005 et 2010, d'après *Europresse*

L'analyse de détail montre que si les occurrences relèvent encore majoritairement du discours rapporté, la prise en charge par le scripteur lui-même est en progression. Le nombre d'occurrences en 2011 semble cependant indiquer que cet emploi de *juste* a atteint un palier.

Pour confirmer ce premier sondage, il convient d'élargir l'enquête. Le point de départ étant les commentaires des locuteurs ordinaires, on a d'abord construit un corpus des remarques sur *juste* : on a mené sur le web les requêtes par le mot cible *juste*, en cooccurrence avec des mots ou séquences du type *expression(s), tic(s) de langage, à la mode*. On a ainsi réuni un ensemble d'articles, de chroniques, de messages postés, qui permettent de recueillir une série de collocations ; plusieurs relèvent la phrase figée *c'est juste pas possible*, et quelques-uns la collocation *juste énorme*, mais, pour le reste, les inventaires sont diversifiés. Si l'on s'en tient aux collocations adjectivales, sont signalés : *juste beau, juste cool, juste hallucinant, juste hilarant, juste génial, juste insupportable, juste pathétique, juste prodigieux, juste sublime, juste superbe, juste terrible, juste trippant*, ainsi que *juste pas croyable, juste pas normal* et *juste trop bien*. Reste à confronter ces signalements avec la réalité de l'usage.

D'après la représentation qu'en donnent les bases de données de presse, ces collocations ne sont pas toutes aussi actives : dans la base *Europresse* interrogée jusqu'au 31/12/2011, *juste génial(es)*, par exemple, présente 99 occurrences, et *juste magnifique(s)* 97 ; en revanche, *juste trop bien*, avec seulement 23 occurrences, se révèle d'usage assez restreint dans l'écrit de presse, de même que *juste hilarant* (8 occurrences) et *juste trippant* (une seule occurrence). Il n'y a pas nécessairement là erreur d'appréciation de la part du remarqueur : l'écrit de presse ne se confond pas avec l'usage oral, généralement visé par les remarques. La confrontation signalement/usage est limitée ici à l'usage écrit de presse.

Dans ce cadre restreint, on peut confirmer la tendance observée pour la collocation *juste incroyable*, comme le graphique ci-dessous le fait apparaître :

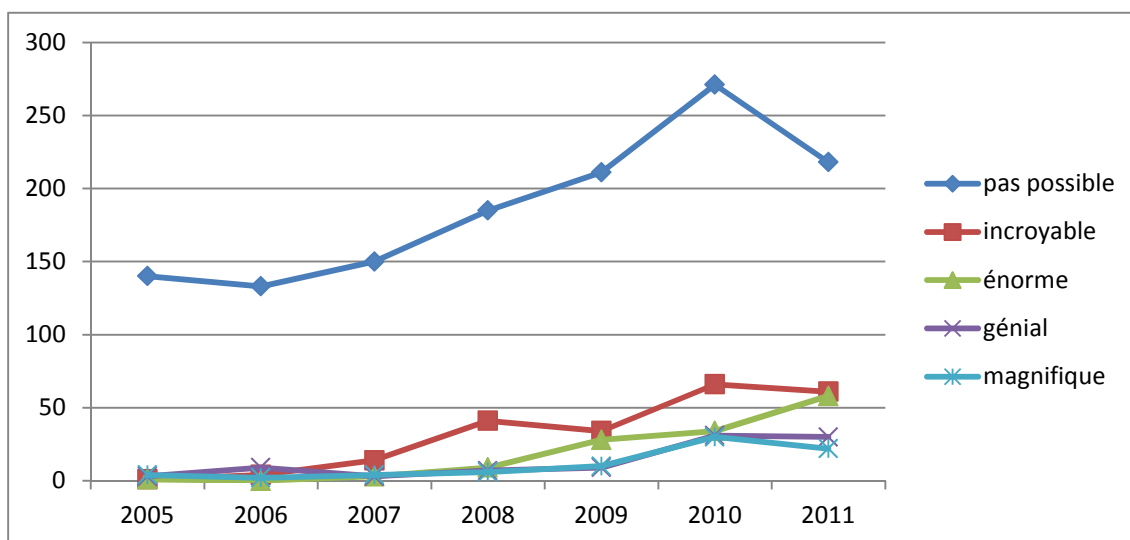


Tableau 2. *Juste pas possible, juste incroyable, juste énorme, juste génial, juste magnifique* entre 2005 et 2010, d'après *Europresse*

*Juste génial, juste magnifique, juste insupportable*, de même que *juste incroyable*, ont atteint leur pic de plus haute fréquence en 2010 et connaissent en 2011 soit une stagnation, soit un déclin. Seule la collocation *juste énorme*, dont le démarrage a été tardif, voit encore sa fréquence croître en 2011. Mais l'affaissement est net pour la phrase figée *c'est juste pas possible*, qui apparaît comme la locomotive de *juste*. Comme *c'est vrai que* en 2004 (Siouffi, Steuckardt, Wionet, 2011), il semble que *juste* soit entré en 2010 dans une nouvelle phase. Son émergence a été repérée et stigmatisée – le remarqueur Frédéric Pommier lui a donné un nom : la « justouille »<sup>4</sup> ; l'empan étroit des études de diachronies courtes ne nous permet pas encore de savoir si l'émergence de *juste* résistera à cette stigmatisation.

Pour le diachronicien, les bases de données de presse présentent ainsi un certain nombre d'avantages, et notamment celui de la fiabilité : elles permettent d'établir des datations et d'identifier les sources. Ces résultats, s'ils sont satisfaisants, doivent cependant être affinés. En effet, l'écrit surveillé des journaux peut dissimuler, ou étouffer des informations que certaines sources moins contrôlées, comme le web, peuvent fournir.

Le web ne correspond certes pas exactement à la définition classique du corpus : « un corpus est une collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques et extralinguistiques explicites pour servir d'échantillon d'emplois déterminés d'une langue » (Sinclair cité par Habert et alii, 1997 : 144). Le web ne présente pas les propriétés d'organisation extralinguistique requises : certaines occurrences ne seront pas conservées du fait de l'instabilité de certains sites, nous ne connaissons rien ou à peu près des scripteurs, les datations précises sont parfois difficiles. Toutefois il n'est plus guère possible à l'analyse linguistique de se priver désormais de cette source de données (Rastier, 2011), particulièrement pour l'étude du français contemporain, car il permet de saisir des managements nouveaux, et, dans une certaine mesure, traçables, de formes éventuellement volatiles, mais parfois émergentes, dans un processus dynamique. Lorsqu'il s'agit d'étudier l'émergence, le web fournit des données en temps réel (Véronis, 2010) : il est nourri quotidiennement des formes linguistiques les plus diverses – français surveillé, français oralisé, régions et âges confondus, etc. Si cette hétérogénéité et cette identification faible des scripteurs constituent une insuffisance lorsqu'il s'agit de mener une enquête sur tel ou tel sociolecte sur des facteurs de variation, elles peuvent au contraire ne pas apparaître comme trop gênantes quand l'objectif est de saisir des émergences à un niveau très général et très partagé de l'usage. Pour autant, la question de savoir si le web constitue un espace plus représentatif d'un état de langue générale que tout autre corpus n'a à notre connaissance pas véritablement été étudiée empiriquement.

Reste, pour l'étude diachronique, un important problème qui est celui de la datation des attestations sur le web. L'internaute ordinaire dispose actuellement d'outils limités : sur le moteur de recherche actuellement dominant, existe, depuis 2008, par « Google recherche avancée » la possibilité de mener une recherche par date. La conception de l'outil n'est malheureusement pas très satisfaisante pour le diachronicien : elle permet seulement de remonter une période maximale de 365 jours précédant l'interrogation ; on apprendra par exemple que, du 19 novembre 2010 au 19 novembre 2011, *juste incroyable* présente 89900 occurrences sur le web, ce qui confirme une installation dans l'usage, mais ne permet pas d'apprécier une évolution. Pour compléter l'investigation, il est donc nécessaire de travailler, en collaboration avec des informaticiens, à affiner les performances des robots d'indexation du web (*crawler*). Dans l'état actuel de la recherche, le repérage des occurrences attestées doit donc croiser deux types d'interrogation : celle du web et celle des corpus et base de données. Si l'exploration de ces terrains donne à observer le processus d'émergence, elle ne nous en livre pas les clefs : l'enquête pourra-t-elle mettre à jour ses mécanismes ?

#### 4 Enquêter sur la diffusion

Si nous reprenons la distinction, aujourd'hui bien reçue en sociolinguistique (voir Milroy 1992 et 1997, par exemple), entre innovation et changement, l'un des intérêts principaux de l'enquête sur des

diachronies courtes et contemporaines devrait être d'approcher de plus près le phénomène de la diffusion, qui permet le changement.

Revenons au préalable sur les principales notions installées dans le champ de la sociolinguistique. On sait que la notion de diffusion a été introduite à partir des propositions de Rogers en 1962. Ce dernier a notamment mis en place une opposition entre ce qu'il appelle l'homophilie dans les relations humaines, type de communication basée sur le partage de ce qui est commun, et que Rogers juge avoir déjà été approchée par le sociologue Gabriel Tarde, dans son ouvrage fondateur *Les lois de l'imitation*, et l'hétérophilie, type de communication basée sur la différence, cruciale selon lui dans la diffusion de l'innovation (Rogers, 2003 : 306).

Dans le second volume de ses *Principles of Language Change* (Labov 2001), W. Labov s'est inspiré des analyses du sociologue Herbert Blumer<sup>5</sup> portant sur la mode, et a jugé que, selon lui, on pouvait observer dans les deux phénomènes une prégnance des dynamiques de la grégarité et de l'influence (Labov 2001 : 361). Il considère ainsi le phénomène du *leadership* dans le domaine de la mode comme un modèle possible pour la représentation du changement linguistique (Labov 2001 : 357)<sup>6</sup>.

De leur côté, L. et J. Milroy distinguent entre trois types d'acteurs, dans le changement linguistique : les *innovators*, les *early adaptors* et les *followers* (voir Milroy et Milroy 1985). Cette décomposition est précieuse pour montrer le rôle que jouent les réseaux sociaux (*social networks*) dans l'apparition et la propagation des changements. Dans ce schéma, Labov a montré le rôle singulier que peuvent parfois jouer ce qu'il appelle les *outliers* (Labov 1997, 1999, 2001) – il désigne par là des individus dotés d'une capacité d'innovation supérieure à la moyenne, et d'un sens différent de la conformité.

Peut-on, sur les diachronies courtes et contemporaines, observer ces phénomènes de leadership et de conformité, que ces travaux fondateurs identifient comme des paramètres essentiels de la diffusion ?

En premier lieu, examinons si la spécificité de l'espace d'énonciation nouveau que constitue le web peut donner accès à des observations locales de micro-dynamiques de diffusion, différentes de celles qui étaient possibles à partir de sources d'écrit traditionnel.

Aux sites classiquement normatifs, comme [www.languefrancaise.net](http://www.languefrancaise.net), ou spécialisés dans les vérifications orthographiques, sémantiques, étymologiques ou grammaticales, sont venus s'ajouter des forums de discussion où s'affrontent purisme et anti-purisme, et surtout des espaces énonciatifs très variés, dont la cartographie, au demeurant sans cesse en mouvement, reste à faire, où se déploient devant les yeux de l'internaute des polémiques autour d'un mot ou d'une expression. Certains sites ou certains blogs donnent facilement accès à ce qui s'apparente parfois à des fils de discussion entamés après qu'un internaute se fut livré à un « lynchage » de mot ou d'expression. Pour l'historien de la langue habitué à l'espace livresque, à la structuration relativement lente (quoique plus rapide que ce que l'on pense ordinairement) construit par ceux qu'on a appelés les « remarqueurs », ces sites offrent comme des raccourcis saisissants des dynamiques discursives qui y étaient parfois observables. L'adoption et le rejet, le goût et le dégoût, la reconnaissance ou le déni d'un marquage, y sont affichés de façon quasi simultanée.

Pour autant, travailler sur ce type d'espaces n'est pas chose aisée. L'idéal serait qu'en suivant un « fil » de discussions, on puisse parvenir à isoler de mini-processus de leadership ou de conformité se mettant en place. Pour reprendre l'exemple du modalisateur *juste*, on a pu relever ainsi qu'un billet posté en septembre 2007 sur le site « Agoravox » (<http://www.agoravox.fr/actualites/societe/article/tout-ce-qui-est-juste-n-est-pas-29224>) avait donné lieu jusqu'à présent à 98 réactions. Pour autant, ces réactions ne semblent pas pouvoir être analysées, pour le moment, autrement que définissant une manière de synchronie d'opinions juxtaposées. Sur Facebook, un mur « C'est juste... : anglicisme insoutenable », a été ouvert août 2010 : il est intéressant de surveiller comment il évolue.

En deuxième lieu, on posera la question – qui émerge fatalement à un moment ou à un autre – de savoir s'il faut prendre en compte tous les lieux de la reprise ludique, détournée, exhibée, transformée que sont l'imitation littéraire<sup>7</sup>, l'humour, la chanson, le cinéma, lieux d'émergence de visibilités particulières. On a assumé plus haut la position générale selon laquelle, la vie du langage ne pouvant être considérée comme uniquement faite d'usages spontanés, mais intégrant la secondarité, le retour critique, la conscience,

l'attitude, le jeu, le pastiche, l'historien de la langue se doit d'associer l'observation de ces événements méta à celle des usages considérés dans leur réalité brute. Il s'agira plutôt ici de se demander si, au sein de ces espaces, apparaissent des phénomènes de conformités autour d'une innovation qui puissent donner lieu à une analyse.

L'une des postulations qui se trouvent en amont de notre recherche, en effet, est que, par certains de ses aspects, dans certains de ses domaines (phonétique, lexical, sémantique, syntaxique), et dans certains de ses contextes, l'innovation linguistique a quelque chose de commun avec le phénomène plus général de la mode. On trouve cette comparaison développée chez Labov (2001 : 360-362, par exemple). Plusieurs linguistes proposent aujourd'hui de considérer la « mode » comme l'une des conditions possibles d'un phénomène émergent. Ainsi Michel Banniard :

La plupart des mutations initiales trouvent une explication solide à partir de la pulsion énonciatrice par laquelle un sujet fait fluctuer la parole commune dont il est dépositaire et acteur pour créer un signal différent. Ce dernier connaîtra le sort de tous les phénomènes de ce type : épiphénomène transitoire tic individuel qui va disparaître avec le locuteur ; succès éphémère qui se propagera sur une zone ou un temps donné ; création réussie qui va, comme on dit maintenant, se contextualiser et se généraliser ; la dernière possibilité est la perpétuation du nouveau signal. Dans ce dernier cas, le diasystème de la langue est modifié. L'accumulation de ces changements conduit au changement de diasystème, à une forme nouvelle de la langue (Banniard et Philips, 2010 : 57).

Considérer des faits linguistiques nouveaux comme des « modes » n'a donc rien en soi de particulièrement dévalorisant et ne devrait pas, même pour le linguiste épris de système, être vu comme une manière de les faire sortir du champ d'une étude diachronique « sérieuse ».

Pour autant, il n'est pas facile de travailler sur ce statut qu'on pourrait qualifier d' « intermédiaire » d'un usage, entre singularité curieuse, encore anecdotique et soumise à renégociation potentielle, et installation réelle dans le paysage de l'idiome. Plusieurs mots ou expressions approchent aujourd'hui ce phénomène, mais aucun n'est réellement satisfaisant.

Le mot de *buzzword*, par exemple, apparu en contexte anglophone dans les années 40 et aujourd'hui employé dans de nombreuses langues, s'appuie sur la reconnaissance du phénomène de la mode<sup>8</sup> comme processus local et rapide de diffusion. La palette sémantique importante qu'on peut dessiner aux usages historiques et actuels du mot est malgré tout là pour témoigner de la variété de dynamiques sociolinguistiques qui peuvent être désignées, englobées, ou sous-entendues par ce terme. Si le vocable désignait à l'origine les « mots-clés » que l'on pouvait extraire d'un texte dans une optique d'apprentissage (« mots balises », en quelque sorte, « mots repères »), ce sens paraît être aujourd'hui tombé en désuétude. Un sens essentiel repéré par les spécialistes de la communication (voir Lafrance 1986) dès les années 80 semble en revanche avoir à faire avec la dynamique sociale de la conformité. Le *buzzword* entendu ainsi est, à l'inverse d'un vocable scientifique, un mot d'extension mal définie propulsé par un univers de discours, politique, économique, commercial ou autre, dans le but d'impressionner, de créer une impression de nouveauté, et de décourager la mise en question. Ce fonctionnement a visiblement un caractère spécifique qui nous paraît sortir de la pure définition des innovations linguistiques destinées à connaître la diffusion. Il pose la question de l'évaluation du fonctionnement de ces mots en *buzzword*, évaluation qui ne peut avoir lieu ailleurs que dans une dimension critique, de pensée critique, par exemple. Considérer les *buzzwords* comme de simples phénomènes émergents faisant tache d'huile reviendrait à leur dénier, précisément, leur caractère de *buzzwords*. Malgré tout, leur étude, à laquelle peuvent servir de balises commodes des repérages et des dénonciations critiques<sup>9</sup>, est sans doute un aspect essentiel de l'étude de la diffusion dans le contexte récent.

Au demeurant, la terminologie à laquelle ressortit *buzzwords* n'est qu'une représentation fragmentaire de tout un univers discursif – impliquant souvent sa mise en question – où l'attestation et le relevé de mots nouveaux s'insèrent dans le cadre d'une interrogation latente, à orientations très diverses, autour du phénomène de la conformité langagière. Les expressions d'*éléments de langage*, de *phrases rituelles*<sup>10</sup>, de



*langue de bois*, de *langue de coton*, de *novlangue* (voir Bihl, 2007) de *politically correct*, de *nouveau wording* etc., rejoignent cette problématique, en se colorant chaque fois d'implications diverses.

Le terme de *buzzwords* appartient au discours de la communication numérique : *curateur*, *data visualisation*, *géolocalisation*, ou encore *cloud computing*, tels seraient, selon les avis recueillis sur le web, les « buzzwords de 2011 ». Celui d'*éléments de langage* a cours dans le monde politique, celui de *nouveau wording* dans l'univers de l'entreprise. Si elle procède de mécanismes analogues, la diffusion des modalisateurs du type *c'est vrai que* ou *juste* se situe à un niveau plus partagé de la circulation des discours.

Cependant, on le sait, les linguistes parlent de *changement* seulement lorsque des signes suffisants existent pour permettre de dire qu'une forme s'est installée dans les usages. À l'évidence, l'observation de micro conformités locales autour d'un usage émergent ne suffit alors pas pour franchir cette étape, condamnant plutôt à parler de « modes ». L'appel aux instances de légitimation et de validation que sont les dictionnaires est-il alors nécessaire ? De fait, un survol même rapide des forums, fils ou blogs tournant autour des faits langagiers nouveaux révèle que la convocation du dictionnaire reste fréquente, l'enregistrement de nouveaux mots dans le *Petit Larousse* ou le *Petit Robert* continuant à faire office de légitimation. Dernière étape sans doute du processus de diffusion, l'apparition d'une nouvelle entrée dans le dictionnaire est aisément repérable, naturellement, mais il n'est pas sûr qu'il faille s'en contenter pour parvenir à une vision exhaustive de l'innovation lexicale entérinée. Le « choix des mots nouveaux » que pratiquent les dictionnaires dans leurs réactualisations est une problématique lexicographique dotée de ses enjeux propres, chaque fois spécifiques, et où la frontière entre terminologie et vision de la langue commune fait l'objet de négociations particulières. À quoi il faut ajouter que les innovations sémantiques ou morpho-syntaxiques, essentielles dans le cas des modalisateurs, apparaissent souvent plus délicates à décrire dans le cadre du dictionnaire, les outils linguistiques fins nécessaires pour le faire étant parfois de maniement peu commode.

On considérera ainsi les instances « traditionnelles » de légitimation comme des acteurs parmi d'autres dans la vaste dynamique qui entoure la propagation plus ou moins large, plus ou moins rapide, plus ou moins éphémère, qui va être celle d'un mot ou d'une forme nouvelle. Ainsi l'enquête sur les usages émergents peut s'arrêter à mi parcours de ce qui sépare l'innovation du changement.

## 5 Conclusion

Enquêter sur de nouveaux usages sur une diachronie courte et sur des corpus écrits apparaît par conséquent comme une opération nécessitant des choix méthodologiques, le principal demeurant d'opter entre une prise en compte des usages attestés à titre de simple matériau linguistique, et une prise en compte simultanée de ce qu'on pourrait dénommer l'*accompagnement* épilinguistique et social d'un usage jugé émergent.

Considérant la langue comme un espace social, on estime que cet accompagnement occupe un statut inaugural : il *institue* à sa manière un usage nouveau ; c'est pourquoi on défend ici la seconde option : c'est de l'événement méta que part l'enquête. Il pointe, pour l'enquêteur, le fait linguistique à observer et il constitue un aspect de la condition langagière des faits émergents

Puisqu'il s'agit d'enquêter sur la diachronie du contemporain, le repérage des commentaires ainsi que des usages commentés ne saurait se limiter à des corpus fermés sur un état de langue révolu. Pour étudier les faits d'émergence dans la langue française d'aujourd'hui, il est nécessaire d'utiliser des bases de données évolutives, et il paraît souhaitable de ne pas s'interdire le recours au web, malgré les réticences que peut légitimement inspirer le caractère instable de ces données. Quelque labiles qu'elles soient, les données du web se laissent cependant saisir, de façon systématique, par les robots d'indexation, ce qui permet d'envisager la création de corpus d'un type nouveau. Munie des outils actuels de fouille du web, l'enquête sur les usages innovants peut, et doit, à notre sens, s'aventurer sur les terrains où se construit et se commente le français d'aujourd'hui.

Il ne suffit cependant pas qu'un usage soit neuf et remarqué pour qu'il devienne émergent : il faut encore qu'il s'installe dans la pratique sociale de la langue. Pour comprendre les différents vecteurs qui opèrent cette installation, il convient à la fois d'explorer bases de données et web, en y cherchant cette fois les lieux de validation des innovations linguistiques.

Pour le moment, ce que font apparaître les premières études que nous avons conduites sur des modalisateurs du dire (*c'est vrai que* et *juste*), cas d'innovations langagières il est vrai particulièrement liées à la parole (dans son sens d'oral comme dans son sens saussurien), c'est que ce dont nous pouvions avoir le pressentiment en nous appuyant sur l'expérience historique des « remarqueurs », à savoir l'intrication des dimensions des usages et des commentaires épilinguistiques ordinaires (ou non) dans les moments d'émergence de faits nouveaux, se révèle vrai. Pour certaines innovations qui impliquent des diffusions rapides, ou des effets de mode, le rôle du « pointage » est important.

Par ailleurs, ces premiers résultats semblent faire apparaître la récurrence d'un « effet de courbe », les phases d'émergence, pour les deux marqueurs que nous avons étudiés de près, étant suivies, d'une façon qui nous a surpris nous-mêmes au cours de l'enquête, d'une décroissance quantitativement assez notable, et, nous semble-t-il, assez précisément datable. Ce phénomène de pure « diachronie » nous paraît un bon indicateur de la pertinence méthodologique – peut-être extrapolable en dehors du français contemporain – des études en « micro-diachronie ». Il amène à porter un nouveau regard sur la distinction théorique passionnante entre innovation et changement.

C'est en permettant de mettre au jour assez précisément ce double caractère que l'étude des diachronies courtes et contemporaines peut présenter, à notre sens, un intérêt et une indéniable spécificité.

## Références bibliographiques

- Achard-Bayle, G. et Paveau, M.-A. (2009). *Linguistique populaire ?*, *Revue Pratiques*, n° 139-140.
- Achard-Bayle, G., Lecolle, M. (2009). *Le sentiment linguistique*, Université de Metz, Ceted, *Recherches linguistiques*, 30.
- Andersen, H. (1988). « Centre and periphery : adoption, diffusion and spread », *Historical dialectology. Regional and Social*, Berlin, New York, Amsterdam: Mouton de Gruyter: 39-83.
- Auroux S. (1986). « Le sujet de la langue : la conception politique de la langue sous l'Ancien Régime et la Révolution », *Les Idéologues. Sémiotique, Philosophie du langage et linguistique pendant la Révolution Française*, W. Busse et J. Trabant (éds), Amsterdam, John Benjamins : 259-276.
- Banniard, M. et Philips, D. (2010). *La fabrique du signe : linguistique de l'émergence entre micro- et macro-structures*, Toulouse, Presses Universitaires du Mirail.
- Berlan, F. (dir.) (2006). *Langue littéraire et changements linguistiques*, Presses Universitaires de Paris Sorbonne.
- Bühr, A. (2007). *La Novlangue néolibérale*, Lausanne, Pages Deux.
- Caron, P. (éd.) (2004). *Les Remarqueurs sur la langue française du 16<sup>e</sup> siècle à nos jours*, Rennes, La Licorne.
- Combettes, B. (2011). « Linguistique diachronique et corpus : problèmes théoriques et méthodologiques », communication au Congrès de l'AFLS, Nancy, ATILF, septembre 2011.
- Galazzi, Enrica, Molinari, Chiara (éds). (2008). *Les français en émergence*, Peter Lang, Bern, Berlin, Bruxelles, Frankfurt am Main, New York, Oxford, Wien.

- Habert, B., Nazarenko, A., Salem, A. (1997). *Les linguistiques de corpus*, Paris, Armand Colin.
- Labov, W. (1997). « Locating the leaders of linguistic change », communication, 71th Annual Meeting of the Linguistic Society of America, Chicago.
- (1999). « The role of outliers in the incrementation of linguistic change », communication 73d Annual Meeting of the Linguistic Society of America, Los Angeles.
- (2001). *Principles of Language Change I : Social factors*, Oxford, Blackwell.
- Lafrance, J.-P. (1986). « L'interactivité : généalogie d'un mot à la mode », *L'espace social de la communication*, Concept et théories, Paris, Retz.
- Lecolle, M. (à paraître), « Sentiment de la langue, sentiment du discours : changement du lexique, phraséologie émergente et air du temps », *Diachroniques* n°2.
- Ligas, P. (2008). « Définition et exemple : quelle complémentarité ? L'illustration du concept dans le Dictionnaire alphabétique et analogique du français des activités physiques et sportives », *Lexicographie et informatique. Bilan et perspectives*, Nancy, ATILF.
- Milroy, J., et Milroy, L. (1985). *Authority in Language. Investigating Language Prescription and Standardisation*, London, Routledge.
- Milroy, L. (1987, 2<sup>nd</sup> ed.). *Language and social Network*, Oxford, Blackwell.
- Milroy, J. (1992). *Linguistic Variation and Change*, Oxford, Blackwell.
- (1997). « Internal vs. External motivations for linguistic change », *Multilingua, Journal of Cross-Cultural and Interlanguage Communication*, Volume 16, Issue 4: 311–324.
- (1999). « Toward a speaker-based account of language change », Jahr, E. (éd.), *Language Change : Advances in Historical Linguistics*, Berlin, Mouton: 21-36.
- , (2003). « On the Discourse of Historical Linguistics : Language-Internal Explanation and Language Ideologies », *Forum Mod Lang Stud*. 2003 39 (Online): 357-370.
- Narjoux, C. (éd.). (2011). « Le français au XXI<sup>e</sup> siècle : continuité et évolution (2) », *L'Information grammaticale* n°130.
- Ollinger, S., Valette, M. (à paraître). « La créativité lexicale : des pratiques sociales aux textes », Actes du 1er Congrès International de néologie des langues romanes (Barcelone, 07 - 10 mai 2008) CINEO'08.
- Paveau, M.-A. et Rosier, L. (2008). *La langue française. Passions et polémiques*, Paris, Vuibert.
- Picton, A. et Condamines, A. (2010). « Vers la définition d'une méthode pour adapter les ressources terminologiques construites à partir de textes », D. Banks (éd.), *Aspects diachroniques du texte de spécialité*, L'Harmattan, Paris : 113-128.
- Rastier, F. (2011). *La mesure et le grain. Sémantique de corpus*, Paris, Champion.
- Rogers, E. (2003[1962]). *Diffusion of Innovations*, 5th. ed., New York, Free Press.
- Siouffi, G., dir., (à paraître), *Modes langagières dans l'histoire. Processus mimétiques et changement linguistique*, Paris, Champion.
- Siouffi, G., Steuckardt, A., Wionet, C. (2011). « La modalisation de l'énoncé à l'écrit : enquêtes sur les représentations et les usages. Le cas de *c'est vrai que* », *L'information grammaticale*, n° 130 : 11-17.
- Tarde, Gabriel ([1890] 2001). *Les lois de l'imitation*, Paris, Les Empêcheurs de tourner en rond.
- Valette, M. (2010). « Méthodes pour la veille lexicale », *Sur les dictionnaires, (Kénitra, le 7 décembre 2007)*, L. Messaoudi, H. El Amrani, B. El Gouak, T. Akkah Afkinich, (éds), Kénitra, *Publication du laboratoire Langage et société*, Université Ibn Tofail : 251-272.

Véronis, J. (2010). « Google : le plus grand corpus linguistique de tous les temps », <http://blog.veronis.fr/2010/12/google-le-plus-grand-corpus.html>

---

<sup>1</sup> On peut penser à Renaud Camus, dans *Répertoire des délicatesses du français contemporain*, Paris, P.O.L., 2000, ou dans *Décivilisation*, Paris, Fayard, 2011, par exemple.

<sup>2</sup> Interrogation faite le 19 novembre 2011.

<sup>3</sup> « L'Italien Max Biaggi est à court de mots. "Je n'ai rien à dire. C'est juste incroyable !" », Dino Dimeo, « Mad Max Biaggi, c'est fou », *Libération*, 6 avril 1998. Pour chaque base, la requête a été menée sur la totalité des archives, le 22 novembre 2011.

<sup>4</sup> Frédéric Pommier, *Mots en toc et formules en tic : petites maladies du parler d'aujourd'hui*, Paris, Points, Le goût des mots, 2011.

<sup>5</sup> Voir Blumer, Herbert, *Symbolic Interactionism: Perspective and Method*, Berkeley, University of California Press, 1969.

<sup>6</sup> Labov note en particulier que les paramètres de l'âge (diminution du leadership avec l'âge) et du sexe (rôle plus grand des femmes) semblent fonctionner de manière comparable dans la mode et dans le changement linguistique (Labov 2001 : 360).

<sup>7</sup> Voir Berlan, éd., 2006.

<sup>8</sup> En français, une des traductions les plus fréquentes proposées à *buzzword* est d'ailleurs « mot à la mode ».

<sup>9</sup> Voir par exemple A. Sokal et J. Bricmont, *Impostures intellectuelles*, Paris, O. Jacob, 1997.

<sup>10</sup> Listées par exemple sur le site Wikipédia de l'émission « Tout le monde en parle », animée par T. Ardisson entre 1998 et 2006, et où le lancement conscient d'innovations langagières faisait partie de l'identité de l'émission.