

ANALYSE D'ÉNONCÉS À FORT IMPACT ET ÉLÉMENTS POUR LEUR DÉTECTION AUTOMATIQUE

Emmanuel QUINTIN

Université de Paris – Sorbonne
emmanuelquintin@hotmail.com

Agata JACKIEWICZ

Université de Paris – Sorbonne, STIH
Agata.Jackiewicz@paris-sorbonne.fr

Thibault ROY

Société NOOPSIS
tr@noopsis.fr

Résumé : L'article présente une méthode d'analyse portant sur des énoncés politiques à fort impact médiatique. Notre recherche vise à terme la construction d'un outil informatique permettant de répondre à deux besoins clairement identifiés aujourd'hui : (i) identifier des discours en circulation, (ii) évaluer et analyser l'impact d'une citation particulière.

Summary : This paper develops a method for analyzing political quotations which draw a lot of attention from the media. It is based on a corpus study of online utterances quoting a well-known French political statement. Our work aims to build up an automated tool, meeting today two needs clearly identified : (i) detect circulation speeches,(ii) appraise et analyse the impact of a peculiar quotation.

Mots clés : discours politique, citation, circulation, impact, méthode d'analyse

Keywords : political speech, quotation, circulation, impact, analysis method

INTRODUCTION

Les nouvelles pratiques discursives propres au Web 2.0 (blogs, wikis, forums, tweets...) diversifient les modes d'emprunt à des discours d'autrui, dont la répétition, plus aisée et plus rapide, multiplie les horizons énonciatifs et les trajectoires argumentatives. Ceci en tension permanente entre le devoir de fidélité à leurs emplois d'origine (Poulard et *al.* 2008) et la recherche d'effets rhétoriques inédits via des transformations parfois substantielles.

À partir d'exemples recueillis sur le net entre avril et juin 2009, nous avons constitué un corpus de soixante énoncés composé de posts, de messages et de billets autour d'une formule devenue célèbre, énoncée en 1990¹ par Michel Rocard : « La France ne peut accueillir toute la misère du monde, mais elle doit savoir en prendre fidèlement sa part ». Nous avons privilégié l'étude de cette citation - qui est la plus fréquemment reprise et offre le plus de variations possibles - par rapport à une autre formule plus ancienne, employée une première fois sur le plateau de l'émission Sept sur Sept du 3/12/1989 et dépourvue de sa structure concessive ; son contenu exact était : « Nous ne pouvons pas héberger toute la misère du monde »². L'énoncé de 1990 a donc infléchi sensiblement le premier.

Krieg-Planque (2009) a pointé les traits définitoires de ce type d'énoncé répété (« immigration clandestine », « pouvoir d'achat », « purification ethnique ») : le caractère figé, la dimension discursive, le fonctionnement comme référent social et l'aspect polémique.

Cette étude entend s'inscrire dans plusieurs traditions. En premier lieu, nous empruntons à l'analyse de discours une approche inscrivant le dialogisme en son cœur ; les énoncés relevés qui constituent notre corpus sont tous dialogiques. Ensuite, nous nous inscrivons dans le sillage des travaux menés à l'ENS-St Cloud par le laboratoire de lexicométrie sur les textes politiques (notamment ceux de Bonnafous et *al.* (1999) sur la circulation des arguments, schèmes et contenus, ainsi que sur les formes de modalités interactives et des représentations du débat public). Enfin, le TAL et notamment les analyses connexes sur le plagiat et le repérage des citations nous ont fourni un support applicatif.

Ce travail suit une démarche itérative, partant de l'analyse des traits caractéristiques d'un énoncé à fort impact. En première approximation, il revient au spécialiste du domaine d'isoler ce type de formule. L'impact est corroboré par des analyses quantitatives et qualitatives (grâce aux formules contextuelles : « déjà culte version » ; « polémique » ; « version qui cartonne » ; « faire le buzz » ; « un tabac » ; « standards copiés ») permettant d'attester de la circulation. Cette notion doit être entendue dans l'acceptation de Rosier (2004) : « La circulation des discours suppose qu'il ne suffit pas d'un simple rapport de discours d'un énonciateur à l'autre mais d'une multiplication de ce rapport : pour être un discours en circulation, un discours doit avoir fait l'objet de plusieurs transmissions ». Ce projet a été réalisé selon deux méthodes différentes : de manière manuelle dans un premier temps et de manière semi-automatique ensuite, en collaboration avec la société Noopsis (Caen).

ANALYSE DE LA FORMULE DE ROCARD

Afin de systématiser l'étude, on a, tout d'abord, conçu une grille de codage, à l'aide d'un tableur. Au sein de cette grille, l'analyse a porté, pour chaque attestation sur :

- (i) la visée pragmatique globale du discours citant ;
- (ii) la forme de discours rapporté, en partant de la typologie de Rosier (2008) ;
- (iii) les expressions introductrices (constructions et lexiques) ;
- (iv) les modifications subies par l'énoncé original.

Il s'agit là de quatre étapes successives de l'analyse, débouchant sur une cinquième :

- (v) l'interprétation des emprunts les moins marqués en surface, autrement dit, de l'effacement énonciatif.

¹ Nous nous sommes intéressés à la formule prononcée lors du cinquantième anniversaire de la Cimade en septembre 1990.

² Ces propos avaient été tenus à la suite de la première affaire de « foulards islamiques », celle du collège de Creil.

1 – LES VISÉES

L'analyse des extraits dans leur contexte a permis d'identifier cinq objectifs pragmatiques, tous dialogiques :

- (i) étayer son propre raisonnement en s'appuyant sur les paroles d'un tiers (11 cas) ;
- (ii) contre-argumenter (condamner l'argument) pour établir son propre discours (15) ;
- (iii) restituer le sens initial de l'énoncé ou rétablir la source (12) ;
- (iv) instaurer une connivence avec un interlocuteur (15) ;
- (v) dévoiler les stratégies énonciatives des réemplois infidèles (7).

1.1 – Étayer son argumentation

Au sens de Maingueneau (2004), la « participation » désigne le régime énonciatif dans lequel des locuteurs collectifs mobilisés peuvent intervenir en communion, tel le slogan politique. Le locuteur citant y montre son adhésion à l'énoncé cité. Dans (1), la reprise pour soutenir son argumentation se fait au prix d'un détournement : étayer revient à contrefaire le sens de la citation originale.

- (1) E.24 : Kouchner a déclaré : « Je me souviens de la phrase de mon Premier ministre socialiste Michel Rocard : "On ne peut pas accueillir toute la misère du monde, mais la France doit en prendre plus que les autres sa part". Eh bien, la France en prendra plus que les autres sa part. »
Source : *Presse web* : <http://fr.novopress.info/?p=8385>

En l'occurrence, Bernard Kouchner – l'énonciateur – se situe dans l'horizon discursif défini par la formule rocardienne. Il est alors membre du gouvernement Sarkozy mais cherche à se réapproprier l'énoncé de son ancien « Premier ministre socialiste Michel Rocard. ». Cet acte d'énonciation est enchâssé dans celui d'un internaute qui outrepassa le propos initial.

Le locuteur peut également s'accorder avec l'énoncé x de l'énonciateur initial en le renforçant par une assertion préliminaire marquée graphiquement (emploi des majuscules) situant son propos dans l'horizon argumentatif. (« je suis ARABOPHILE »). Ce dispositif discursif est proche de celui de la concession mais s'en différencie car on ne peut guère en rétablir la structure prototypique (w, bien sûr, mais Y) telle que définie par (Bres, 1999)

- (2) E.28 : je suis ARABOPHILE mais la France ne peut pas accueillir toute la misère du monde. et la France est un pays Catholique, qui doit le rester
Source : *Forum* : www.boursorama.com/forum/message.phtml?id_message=386928886

1.2 - Condamner l'argument

Selon Maingueneau (2004, p. 119), « l'énonciation du slogan militant implique l'existence d'un extérieur hostile ou indifférent par rapport auquel se pose le groupe ». On relève certains traits spécifiques à notre corpus comme la présence de guillemets, mais la citation, si elle est notée, n'est pas attribuée. À un NOUS, s'oppose un complémentaire ON sujet de l'énoncé à valeur de ILS, placé dans le même champ discursif.

A proprement parler, la « citation » imputée à Rocard, sous sa forme concessive, n'est pas un slogan, mais transformée sous cette forme : ci-dessous, E. 14 assigne à un énonciateur aux relents xénophobes (désigné par « on ») un slogan issu, par détournement, d'un énoncé imputé à Rocard. Toutefois, rien ne prouve que l'énonciateur de E.14 soit conscient de ce double dialogisme (ni le contraire, d'ailleurs) : le « on » écrase la « strate Rocard » obérant un horizon énonciatif.

- (3) E.14 : *"On ne peut pas accueillir toute la misère du monde" nous répète-t-on. Les pauvres sont trop nombreux dans le monde, nous (pays riches) n'y pouvons rien. On nous dit même qu'on ne peut plus aider « nos » propres pauvres (et qu'on doit couper dans les RMIs, services publics et autres formes de la Sécurité Sociale).*
Source : *Blog* : <http://blog.kilobug.org/blog/index.php/2004/08/07/4-on-ne-peut-pas-accueillir-toute-la-misere-du-monde-mais-peut-on-la-resorber>

1.3 - Restituer le sens initial

Dans onze occurrences, on restitue le second membre, très fréquemment omis, de la citation ; un emploi assigne la citation à son auteur.

- (4) E.2 : *Souvenez-vous, Michel Rocard avait prononcé cette phrase restée célèbre sur l'immigration : « La France ne peut pas accueillir toute la misère du monde... » C'était en fait tronquer sa phrase de sa moitié et du sens initial que lui donnait l'ancien premier-ministre de François Mitterrand : Michel Rocard : "La France ne peut pas accueillir toute la misère du monde, raison de plus pour traiter correctement la part qu'elle se doit d'en prendre."*
Source : Blog : <http://www.rmc.fr/blogs/lesgrandesgueules.php?post/2008/09/15/Michel-Rocard>

1.4 - Jouer sur la connivence avec le lecteur

Selon Salvan (2009, p.101), « lorsque l'allusion fait entendre la voix d'un énonciateur collectif ou proverbial, la connivence recherchée est plutôt culturelle et postule un énonciateur partageant les mêmes références, qui sait reconnaître la voix de la doxa, de la phraséologie chrétienne, littéraire, sociale, etc. »

Le premier type de détournement est celui opéré dans la situation d'énonciation :

- (5) E.37 : *Comme par miracle, le président gabonais Omar Bongo vient d'en découvrir une soixantaine chez lui et s'apprêterait à les expulser. Étonnant quand on sait qu'en France l'immigration est assimilée à la misère, d'où cette célébrissime phrase de l'ex Premier ministre socialiste Michel Rocard, devenue le credo de tous les hommes politiques français : « La France ne peut pas accueillir toute la misère du monde ». Ainsi donc la très prospère France aurait un surplus de misère qu'elle s'ingénierait à déverser sur la déjà miséreuse Afrique ?*
Source : Blog : <http://www.afriquechos.ch/spip.php?article3071>

Le segment « *devenue le credo* » manifeste explicitement la transformation en slogan de l'énoncé.

Le second type de détournement s'accomplit dans l'énoncé lui-même. Il s'agit encore d'une pratique faisant appel à la connaissance du lecteur :

- (6) E.40 : *Certes, Nicolas Sarkozy hésite : en dépit des promesses, la France ne peut pas accueillir toute la misère féminine du monde.*
Source : Blog : <http://www.regards.fr/article/?id=3056>

L'énonciateur, par une forme de DDL, attribue dialogiquement à N. Sarkozy l'assertion rocardienne. En l'absence d'un cotexte plus large, on ne peut assurer qu'il fasse allusion à l'énonciation de Rocard : il peut aussi bien ignorer cette source discursive. Mais, peu importe d'une certaine manière car ce slogan, même transformé, fait signe au lecteur.

1.5 - Dévoiler les stratégies énonciatives

Les locuteurs peuvent également s'intéresser à la visée pragmatique, celle-ci constituant alors le message même de l'énoncé. La citation de Rocard est devenue un stéréotype, un topos par rapport auquel les locuteurs se situent.

- (7) E.11 : « *Les immigrés sont tous des sans-papiers et ne veulent pas travailler.* » « *Nous ne pouvons pas accueillir toute la misère du monde* ». *Cette phrase est devenue une expression de sens commun ; elle n'appartient plus à son auteur, elle fait partie des tics de langage, pour l'infirmer ou la récuser, de tous ceux qui ont quelque chose à dire sur l'immigration.*
Source : Blog : http://www.lecavalierbleu.com/images/30/extrait_147.pdf

Le locuteur exhibe le travail de légitimation construit par l'utilisation de la formule ; d'autre part, usant du discours direct entre guillemets, puis de l'anaphore « *cette phrase* », l'énonciateur s'efforce de restituer la citation ; il ne montre aucunement qu'il la reformule ; il la restitue, même si ce n'est pas littéralement exact.

- (8) E.29 : « *La France ne saurait accueillir toute la misère du monde, raison de plus pour qu'elle traite bien la part qu'elle ne saurait manquer d'en prendre.* » *Je ne vais pas me prononcer sur l'auteur de cette phrase et sur son intention. Lui seul sait ce qu'il voulait dire et dans quel but... Par contre, il est*

évident que tous les esprits « raisonnables », « réalistes », associés, parfois à leur corps défendant, avec tout ce que la France compte de racistes, tous ces esprits se sont rapidement approprié la première moitié de la citation. Et pour cause ! Quelle merveille d'hypocrisie !
 Source : Blog : <http://www.cairn.info/revue-rue-descartes-2002-3-page-31.htm>

2 – FORMES DE DISCOURS RAPPORTÉ

En deuxième lieu, les formes de discours rapporté ont été étudiées selon les analyses en continuum de Laurence Rosier (2008). La citation est rapportée principalement (38/60) au discours direct. Les autres formes se distribuent de manière pratiquement équivalente (DDL : 6, DI avec îlot énonciatif : 7, Formes aux confins : 7, DI : 2).

2.1 - Corrélations entre modes de DR et visées pragmatiques

Nous avons tenté de dégager des corrélations entre le mode de DR employé et la visée pragmatique.

	DD	DI	DDL	DI avec IE	Formes aux confins	TOTAL
DÉTOURNER	6		2	1	2	11
CONDAMNER	9	1	1	4		15
RESTITUER	11			1		12
DÉVOILER	5	1	1			7
JOUER	7		2	1	5	15
TOTAL	38	2	6	7	7	60

Tableau 1 : Modes de discours rapporté et visées pragmatiques

Les visées du dévoilement et surtout de la restitution/attribution utilisent essentiellement le DD. Celle de la connivence emploie d'autres formes, même si elle y recourt de manière surprenante souvent. Les formes aux confins sont utilisées pour instaurer une connivence avec le lecteur et, à un degré moindre, pour étayer le propos. L'instauration de la connivence a lieu au DD ou par les formes aux confins.

2.2 - Visées et sur/sous-énonciation

Ces visées correspondent à différents positionnements du locuteur. Celui-ci choisit de « parler avec » lorsqu'il s'appuie sur la formule de Rocard pour la détourner, quand il dévoile les stratégies énonciatives à l'œuvre dans son usage, ou lorsqu'il joue sur la connivence du lecteur. À l'inverse, il choisit de « parler de » lorsqu'il condamne ou restitue la citation.

3 - LA CONSTRUCTION D'UNE MICRO-GRAMMAIRE

Dans un troisième temps, on s'est attaché à construire une grammaire des expressions introductrices de cet énoncé, à partir des segments tels que : « citation très souvent utilisée, partiellement, mutilée », « cette phrase devenue une expression de sens commun », « phrase constamment réduite à... ». Ces expressions de contextualisation (cf. Krieg-Planque) ont été analysées en fonction du mode de citation.

Modes de citation	Total (37)
Discours direct	27
Discours direct libre	2
Discours indirect	2
Discours indirect avec IE	4
Formes aux confins	2

Tableau 2 : Formes de DR dans les phrases introductives

Il est à noter que 20 emplois (sur 37) témoignent explicitement du phénomène de circulation.

3.1 - Corrélations

Plus la forme de DR est canonique, plus on détecte des formules contextuelles, notamment des formules qui insistent sur le succès de cette citation (« cette célébrissime phrase... devenue le credo de tous les hommes politiques », « cette déclaration, restée à juste titre fameuse »). De même, inversement, on constate une présence importante de ces formules contextuelles dans les reprises qui visent à contre-argumenter et à dévoiler les stratégies argumentatives des reprises infidèles. Cet emploi est relativement plus fréquent : 5 cas sur 7. Pour détecter des formules contextuelles et constituer un lexique, le repérage pourrait porter sur ces cas d'emplois.

3.2 - Constitution de patrons morpho-syntaxiques

À partir des marques de surface, on a constitué des patrons morpho-syntaxiques. Ces formules s'organisent autour de substantifs : « citation », « phrase », « formule » auxquels viennent s'ajouter des éléments auxiliaires : les adjectifs et les adverbes, antéposés ou postposés.

(9) E.59 : « *La France ne peut pas accueillir toute la misère du monde* ». Cette phrase percutante dite par Michel Rocard est devenu un credo à droite pour promouvoir la politique du ministre de l'intérieur Mr Sarkozy.

Source : Blog : <http://www.politique.com/forum/showthread.php?t=351>

Une attention spécifique a été portée aux verbes d'énonciation « reprendre » (« reprendre l'idée à son compte »), « citer » (« on a beaucoup cité la phrase »), « répéter » (« nous répète-t'on ») qui, par leur sémantisme, témoignent du phénomène de circulation.

ADJ.	SUBS.	(ADV.)*	ADJ.	ADV.
célébrissime	citation	très souvent	utilisée mutilée apocryphe	partiellement
	phrase	constamment	réduite percutante	
		trop	célèbre	
	déclaration	(restée à juste titre)	fameuse	

Tableau 3 : Quelques patrons morpho-syntaxiques des formules contextuelles

4 - ÉTUDE DES VARIATIONS DE L'ÉNONCÉ

La quatrième étape de l'analyse est celle qui cherche à systématiser les transformations subies par l'énoncé original. L'omission récurrente de son second membre (dans 35 cas sur 60) est la plus manifeste d'entre elles. Seuls huit énoncés rapportent fidèlement la citation originale. Les énoncés fidèles comportent dans cinq cas trois niveaux énonciatifs ou même plus. La fidélité de l'énoncé n'est donc pas corrélée au nombre d'horizons énonciatifs.

On peut distinguer différents degrés d'importance des éléments syntaxiques (§4.1 - 4.5).

4.1 - Les modifications adverbiales : ajout de pas (31), omission de fidèlement (48)

La modification la plus notable par rapport à la citation originale, c'est l'ajout de la particule « pas ». Seules douze phrases s'affranchissent de la tournure adverbiale classique « ne... pas ». De même, l'adverbe « fidèlement » est absent de toutes les reprises infidèles, à l'exception de quatre énoncés.

4.2 - La substitution du sujet (on, nous, la Belgique...) (14)

De nombreux énoncés reformulent la citation originale : la France est remplacée par le pronom « on » (six cas), « nous » (deux exemples) – ce qui conserve à peu près la référence – ou remplacée par un autre nom propre, qui thématise un autre référent (« L'Italie »). Certains énoncés peuvent être à la forme impersonnelle (« il fallait » ou « Peut-on » à la forme interrogative). Le changement de sujet affecte parfois le syntagme verbal et ses éléments : le syntagme « sa part » présent dans la plupart des énoncés (quatre cas) est modifié en fonction du sujet : (« notre part » : trois cas).

4.3 - La variation des verbes modaux (savoir, pouvoir, devoir) (13)

Le syntagme verbal *peut* est également modifié : il se présente à une autre forme temporelle (« pouvait ») ou avec un changement de modalisation (« n'a pas à » ; « n'a pas vocation à » ; « est dans l'incapacité d' »). Dans le second membre, le syntagme verbal composé « doit savoir en prendre » peut parfois être abrégé : « doit/devons en prendre » (cinq cas) ou reformulé (« doit en assumer »). Il peut être remplacé également par un autre syntagme verbal complexe : « peut faire ». La proximité phonologique *avons/devons/savons* autorise également la substitution.

4.4 - Le remplacement des verbes (héberger, importer, en assumer) (7)

Le verbe à l'infinitif *accueillir* peut être remplacé par d'autres : *héberger*, *importer*, *prendre en charge*, mais il reste prépondérant. Il demeure le pivot de la citation originale. La présence de ce syntagme provoque alors nécessairement la présence de l'autre élément « toute la misère ». Les nœuds « accueillir » et « toute la misère » sont bien le cœur du modèle : tous les énoncés comportent ces deux éléments. Bien plus, certains énoncés qui avaient dérogé à la citation originale et notamment à l'élément (« accueillir ») aboutissent alors à « toute la misère ». Rappelons dans le sillage des analyses de Salem (1987) que les techniques de lexicométrie sont particulièrement bien adaptées pour repérer des associations de mots répétés : expressions figées, lexies, complexes, discours répété, phraséologie... comme le segment « toute la misère du monde ».

4.5 - L'insertion d'adjectifs (misère salariale, féminine, intellectuelle) (5)

Seuls quelques énoncés échappent au modèle décrit précédemment : certaines phrases reformulent la citation originale. Elles ajoutent soit un adjectif à l'énoncé original : « salariale », « intellectuelle » ou « féminine » (40), soit un syntagme nominal, « de l'Afrique ».

4.6 - Détournement et effacement énonciatif

Enfin, nous avons considéré les cas des reprises peu ou pas marquées graphiquement (14/60). Plus on fait appel à la connivence du lecteur (savoir partagé), moins on lui fait signe. Mais, par ailleurs, moins on explicite l'acte d'emprunt, plus on s'approprie le propos, et donc plus celui-ci échappe à son auteur. Par là, nous entendons l'effacement énonciatif au sens de Rabatel (2004).

(10)E.38 : *Ce n'est pas une raison pour que la France accueille toute la misère intellectuelle du monde. Surtout qu'en la matière, elle a déjà plus que sa part, jusqu'au plus haut niveau de l'État.*
Source : Blog : http://www.e-torpedo.net/article.php3?id_article=1149

Dans cette tension entre figement et dédoublement énonciatif, sont à l'œuvre des mécanismes dialogiques et de légitimation. Les termes diffèrent : Sarah Leroy (2005) évoque un processus de sédimentation et Laurence Rosier (2004), le dispositif de la métonymie énonciative.

5 - CONCLUSION

Cette première étude a permis l'observation de phénomènes caractéristiques des reprises sur le web des énoncés à fort impact : un nombre limité d'objectifs communicationnels, l'omniprésence des

³ La déclinaison du discours du devoir est classique dans le discours politique, elle prend chez Chirac un tour systématique que confirme la statistique : se présentant comme un conseiller et un censeur, Chirac ne cesse d'indiquer « ce qui doit ou devrait être fait » (Mayaffre, 2004, p. 171).

formes traditionnelles de discours rapporté, leur fréquente transformation, un jeu entre sous et sur-énonciation et surtout la présence récurrente de formules contextuelles autour de la citation. Ces formules environnantes témoignent, dans plus d'un emploi sur deux, de la circulation discursive. Nous avons fait l'hypothèse que ces formules permettent de repérer d'autres énoncés à fort impact.

EXPÉRIMENTATIONS INFORMATIQUES

Afin de tester la pertinence de ces observations, on a tout d'abord recherché, dans des articles de presse, des billets de blogs, etc. présents sur Internet, d'autres formules en circulation. De cette façon, on a ainsi pu constater la fréquence de reprises de certains énoncés :

(11) « *Vous n'avez pas le monopole du cœur* » (Valéry Giscard d'Estaing)

(12) « *Casse-toi pôv'con* » (Nicolas Sarkozy)

(13) « *Un ministre, ça ferme sa gueule, si ça veut l'ouvrir, ça démissionne* » (Jean-Pierre Chevènement)

(14) « *Quand il y en a un, ça va. C'est quand il y en a beaucoup qu'il y a des problèmes.* » (Brice Hortefeux)

On a ensuite cherché à repérer des reprises de ces citations dans un corpus de travail constitué de 500 blogs portant sur l'actualité sociale et politique (des principaux hébergeurs généralistes - *Wikio*, *CanalBlog* - ou spécialisés - *Libération*, *Le Monde*...). Isoler de telles reprises pourrait ainsi permettre de détecter de nouvelles formules contextuelles, ces dernières permettant alors de déceler de nouvelles citations à fort impact. La détection de ces reprises de citations célèbres a été réalisée de la façon la plus automatique possible, à l'aide d'un outil informatique de traitement automatique des langues. L'objectif de cet outil informatique est d'assister son utilisateur dans la détection de reprises de citations célèbres et, ainsi, de l'aider, notamment à identifier de nouvelles formules contextuelles. Pour mettre en place un tel outil, on a utilisé la plate-forme *SemioLabs* développée par la société *Noopsis* (<http://www.noopsis.fr>). Cette plate-forme permet de créer et d'appliquer de façon successive différents analyseurs textuels, tels des tokeniseurs⁴, des étiqueteurs, etc.

6 – TRAITEMENTS INFORMATIQUES

L'outil réalisé avec la plate-forme *SemioLabs* prend donc en entrée le texte à analyser et la citation étudiée. Ensuite, l'outil recherche dans le texte des passages « proches » de la citation d'origine. Afin de pouvoir comparer ces deux segments, nous avons utilisé une distance d'édition, en nous inspirant de Brunet (2003) et de Levenshtein (1965). La distance d'édition est appliquée ici sur les lemmes, elle revient donc à compter le nombre de lemmes qu'il faut supprimer, insérer et/ou remplacer pour passer d'une séquence à l'autre. Ce calcul de distance s'effectue au sein d'un virgule⁵. Nos citations d'origine sont ainsi constituées d'un ou plusieurs virgules, couvrant une partie de phrase, une phrase ou même plusieurs phrases. Cette procédure succède aux étapes classiques de découpage et lemmatisation du texte. Afin d'identifier les reprises les plus proches des citations originales, nous avons sélectionné celles dont la distance est inférieure à certains seuils que nous avons fixés empiriquement sur la base d'un travail d'évaluation des 100 meilleures reprises ramenées pour chacune de nos citations. Cette sélection, empirique et arbitraire, permet d'observer des variations de la formule originale tout en étant encore assez proche de la formule originale.

7 - REPRISES AVEC DES VARIANTES

Sur la base des traitements informatiques présentés précédemment appliqués à notre corpus de travail (500 blogs), nous avons obtenu un grand nombre de reprises. Ces reprises sont tout d'abord analysées dans leurs variantes (cf. section 4).

⁴ Un tokeniseur permet de segmenter les éléments (les mots) du flux textuel grâce aux délimiteurs (tabulations, espaces, renvois à la ligne)..

⁵ Au sens de (Lucas, 2001), un virgule est un constituant délimité par deux ponctuations.

Citation d'origine : « Vous n'avez pas le monopole du coeur »	
« Vous n'avez pas le monopole du socialisme, ni de la raison, ni du coeur. »	
« Vous n'avez pas le monopole du care »	
« Vous n'avez pas le monopole de la souffrance »	
« Vous n'avez pas le monopole du coeur, mais les Dauphinois ont le monopole du gratin »	
Citation d'origine : « Casse-toi, pauvre con »	
« Casse-toi pauvre filaplomb »	
« Casse toi fils de pute d'hongrois »	
« Casse-toi de mon petit écran »	
Citation d'origine : « Quand il y a un ça va... c'est quand il y en a beaucoup qu'il y a des problèmes »	
« Quand il y un raciste, ça va, c'est quand il y en a beaucoup qu'il y a des problèmes »	
« un préfet ça va, c'est il y a plusieurs racistes qu'il y un problème »	
« Une comptable, ca va, c'est quand il y en a beaucoup que cela pose problèmes »	
« Un cigare payé par le contribuable pour un élu, ça va. c'est quand il y en a beaucoup qu'il y a des problèmes »	
Citation d'origine : « Un ministre, ça ferme sa gueule, si ça veut l'ouvrir, ça démissionne »	
« Un parti, ça ferme sa gueule, si ça veut l'ouvrir, ça se présente aux élections. »	
« Un inspecteur ça démissionne ou sa ferme sa gueule »	

Tableau 4 : Variantes des citations originelles

8 - DES REPRISES ... AUX NOUVELLES FORMULES CONTEXTUELLES

Chacune de ces quatre citations a permis de collecter semi-automatiquement – le corpus est constitué automatiquement mais le repérage est manuel – des formules contextuelles (12 pour la citation giscardienne, 8 pour la « pépite sarkozienne », 12 pour celle d'Hortefeux et 10 pour la petite phrase de Chevènement), dont certaines sont reproduites dans le tableau 5.

Vous n'avez pas le monopole du coeur	« réplique de » ; « réplique superbe » ; «réplique qui passera à la postérité» ; «célèbre réplique» ; «phrase qui restera dans la mémoire collective» ; «phrase culte» ; «le fameux»
Casse-toi pauvre con	« déjà culte version » ; «polémique» ; « version qui cartonne » ; «faire le buzz» ; «un tabac» ; «standards copiés»

Quand il y en a un, ça va...	« tohu-bohu médiatique » ; «la bonne blague» ; « le scandale du jour» ; « la blague douteuse» ; « la polémique autour de cette simple phrase» ; « Polémique autour d'un dérapage verbal» ; « vive polémique» ; « la mauvaise blague » ; « scène qui fait le tour de la presse» ; « nouvelle "affaire"»
Un ministre, ça ferme sa gueule, si ça veut l'ouvrir, ça démissionne	«une expression (que nous a) léguée...» ; « ces deux phrases, très célèbres » ; « l'adage » ; « cet adage, la formule popularisée » ; « sa petite phrase », « célèbre phrase », « formule qui a eu un certain succès »

Tableau 5 : Formules contextuelles obtenues

À l'aide de ces nouvelles formules contextuelles détectées, on a tenté de déceler de nouvelles citations en circulation et, ainsi, de valider notre démarche itérative. Nous avons donc injecté trois de ces formules (en gras dans le tableau ci-dessus et soulignées au sein des énoncés) et avons constaté avec satisfaction que nous obtenions des résultats confirmant le bien-fondé de notre approche. Le premier exemple est particulièrement intéressant car il comporte également une formule contextuelle impliquant le phénomène de circulation (« conclusion rabâchée en boucle »).

(15) *S'il n'y avait 228 passagers du Vol Af 447 assurant la liaison Rio-Paris présumés morts, le tohu-bohu médiatique, depuis lundi matin, sur leur disparition passerait pour un gag. (...) L'intensité des commentaires laisse en effet pantois : 48 heures d'explications pour une conclusion rabâchée en boucle : « On ne sait rien ! » Nada !*

(16) *Si, selon l'adage, les grandes équipes ne meurent jamais, alors Biarritz est de celles-ci. (Sports.fr, 5/9/2010)*

(17) *C'est une petite phrase qui risque fort d'énerver, aujourd'hui les responsables de l'information de TF1. Elle est signée d'Alain Vautier, le nouveau Directeur des programmes de France 2 dans la Lettre de l'Audiovisuel. Il affirme : « Nous sommes copiés par TF1, en particulier sur le journal de Laurent Delahousse, le week-end à la mi-journée. TF1 en a repris les bonnes idées. Mais ce n'est pas grave, mieux vaut être copié que copieur (...). » (www.jeanmarcmorandini.com, 6/9/2010)*

Ainsi, de manière itérative et cumulative, sur un corpus déterminé :

- (i) on construit et enrichit la grammaire des expressions introductrices et le lexique associé,
- (ii) on constitue un stock de formules en circulation.

9 – DISCUSSION

Les variantes des énoncés à fort impact sont plus intéressantes pour notre recherche que les reprises fidèles. Elles comportent plus de formules contextuelles : 25 % des variantes comportent une formule contextuelle contre 10 % des reprises fidèles, même s'il faut relativiser ces résultats en fonction de la taille du corpus analysé. Sur le plan qualitatif, la popularité d'une « petite phrase » peut tenir à de nombreux aspects impossibles à tester ici : son pouvoir évocateur, sa facilité de transposition humoristique, surtout dans le cas de celle d'Hortefeux, le caractère illocutoire (les insultes à la « casse toi pauvre con » sont sommaires mais difficilement transposables). Il y a sur l'échantillon limité arbitrairement au seuil d'une centaine d'énoncés, ainsi une vingtaine de reformulations des citations : « Vous n'avez pas le monopole du cœur » et « Quand il y en a un, ça va... » contre une à trois pour « Casse-toi pauvre con » et « Un ministre, ça ferme sa gueule, si ça veut l'ouvrir, ça démissionne ».

D'autre part, compte tenu des différences sensibles sur le nombre d'atomes de citation⁶, il est très difficile de comparer une à une les citations qui ont des comportements spécifiques ; cependant, nous avons repéré quelques tendances. Le taux de formules contextuelles – sur un corpus, encore une fois, limité – dépasse les 10 % : (12 % pour « Vous n'avez pas le monopole du cœur » ; 8% pour « Casse-

⁶ Les atomes de citation désignent les termes de la citation particulièrement porteurs de sens, les « mots-pleins »

toi pauvre con » ; 12 % pour « Quand il y en a un, ça va... » ; 10 % pour « Un ministre, ça ferme sa gueule... ». Compte tenu de la taille du web, la précision est impossible à calculer. Mais le taux de rappel⁷ varie très sensiblement : 90-95 % pour les citations les plus récentes (97 % pour « Quand il y en a un, ça va... » ; 90 % pour « Casse-toi pauvre con ») et 50 % ou moins encore pour celles plus anciennes (51 % pour « Vous n'avez pas le monopole du cœur » et 38 % pour « Un ministre, ça ferme sa gueule... »).

Le point satisfaisant demeure quant à la méthodologie suivie : la démarche itérative, même si elle demande à être affinée, obtient des résultats encourageants. Il conviendrait certainement d'utiliser des méthodes quantitatives plus approfondies pour améliorer notre seuillage⁸, étape indispensable pour automatiser le processus de détection de formules contextuelles et d'énoncés à fort impact.

Enfin, la circulation diversifie les horizons énonciatifs. Si certaines reprises modifient, altèrent ou détournent le sens, d'autres multiplient les strates énonciatives. Quelqu'un dit à l'internaute que quelqu'un a dit quelque chose.

(18) E.31 : *Lors de l'émission de Canal+, Grand Journal, la première dame de France Carla Bruni a eu du mal à s'exprimer sur la pauvreté et ses solutions. Du coup, elle a utilisé une citation de Michel Rocard: « On ne peut pas accueillir toute la misère du monde... ». Seulement, la citation avait une suite: « ...mais nous devons en prendre notre part. » Carla s'était retrouvée un peu embrouillée et Michel Denisot a mis fin au débat... heureusement !*

Source : Info : <http://www.1001actus.com/carla-bruni-la-pauvrete-nest-pas-son-fort-3191.html>

On a pu distinguer, à l'instar des nouvelles pratiques sur le web, une superposition des couches énonciatives : ainsi, on a pu dégager dans une série d'échanges par post jusqu'à quatre horizons énonciatifs.

10 - CONCLUSION ET PERSPECTIVES

Interroger les nouvelles pratiques énonciatives sur le web autour des énoncés répétés consiste, dans l'approche que nous proposons, à étudier les objectifs communicationnels et pointer les horizons énonciatifs. Ce questionnement met en relief également le figement et touche donc à la phraséologie, puisque notre étude porte sur le discours politique. Le web constitue ainsi un laboratoire des fonctionnements discursifs et permet d'observer l'évolution dynamique (échanges d'opinion) et statique (reprise d'énoncés) des représentations culturelles : le processus de stéréotypage au sens d'Amossy (2010). Les nouveaux médias sont en effet très fortement doxiques. On s'appuie sur ce trait pour mettre en place une démarche itérative de repérage de nouvelles citations à fort impact.

Outre cet objectif, l'étude peut avoir comme prolongement l'évaluation du degré de figement du discours politique. Les études sur la stéréotypie et la « langue de bois » sont déjà nombreuses, citons (Pineira et Tournier 1990) entre autres. Adéquatement outillées et menées sur des corpus actuels, en grand volume, et construits selon des paramètres précis, elles pourront livrer une image plus précise et plus complète des mécanismes à œuvre. Une autre piste permettant de prolonger cette étude serait de mesurer l'impact réel d'un personnage politique à travers celui de ses énoncés. Ces études offriraient l'avantage de pouvoir exploiter à la fois les méthodes qualitatives et celles quantitatives.

BIBLIOGRAPHIE

- Amossy, R (2010)., *L'argumentation dans le discours*, Coll. Cursus, Armand Colin, Paris.
Bonnafous, S. & D. Demarchelier (1999), « Quand les députés coupent le 'RESEDA' », *Mots*, n°60, pp. 93-109.
Bres, J., Haillet P., Mellet, S., Nølke, H., Rosier, L., (éds.) (2005), *Dialogisme, polyphonie : approches linguistiques*, Duculot, Paris.

⁷ Le rappel est défini par le nombre de documents pertinents retrouvés au regard du nombre de documents pertinents que possède la base documentaire, en l'occurrence, le web.

⁸ Le seuillage consiste à délimiter un ensemble afin d'optimiser son observation.

- Bres, J. (1999), « Vous les entendez ? Analyse du discours et dialogisme », *Modèles linguistiques*, XX, 2, 1999.
- Brunet, E. (2003), « *Peut-on mesurer la distance entre deux textes ?* », *Corpus*, n°2, La distance intertextuelle.
- Krieg-Planque, S. (2009), *La notion de formule en analyse de discours, Cadre théorique et méthodologique*, Presses universitaires de Franche-Comté, Besançon.
- Leroy, S. (2010), « Contrats et emprunts entre discours : l'exemple du discours politique chez Deleuze », *Congrès mondial de Linguistique française*, La Nouvelle-Orléans, Etats-Unis, 12-15 Juillet 2010.
- Leroy, S. (2005), « Le détournement dans les titres de presse : un marquage dialogique ? », in Bres Jacques et al. (dir.), *Dialogisme et polyphonie*, Bruxelles, De Boeck-Duculot, pp. 201-214.
- Levenshtein A. (1965), « Binary codes capable of correcting spurious insertions and deletions of ones », *Problems of Information Transmission*. Vol. 1, Issue 1, pp. 8-17.
- Lucas, N., Giguet E., Vergne J.,(2001), « Détection automatique de la citation et du discours rapporté dans les textes informatifs ». *Le discours rapporté dans tous ses états : question de frontières*. Bruxelles.
- Maingueneau, D. (2004) « Hyperénonciateur et 'participation' », *Langages*, n°156, pp. 111-126.
- Mayaffre, D. (2004), *Paroles de président. Jacques Chirac (1995-2003) et le discours présidentiel sous la Vème République*, Honoré Champion,
- Pineira, C., Tournier M. (1990), « *De quel bois se chauffe-t-on? Origines et contextes actuels de l'expression langue de bois dans la presse française* », *Mots* n°21, pp. 5-20.
- Poulard F., Waszak T., Hernandez N., Bellot P. (2008), « *Repérage de citations, classification des styles de discours rapporté et identification des constituants citationnels en écrits journalistiques* ». *Actes de TALN*, Avignon, France.
- Rabatel, A. (ed) (2004), « *Effacement énonciatif et discours rapportés* », *Langages*, n°156, pp. 3-17.
- Rosier, L. (2008), *Le discours rapporté en français*, Ophrys, Paris.
- Rosier, L. (2004), « La circulation des discours à la lumière de «l'effacement énonciatif» : l'exemple du discours puriste sur la langue », *Langages*, n°156, pp. 65-78.
- Salem, A. (1987), *Pratique des segments répétés*, Publications de l'INaLF, collection "St.Cloud", Klincksieck, Paris.
- Salvan, G., « Réécrire de connivence : les fortunes dialogiques de l'allusion », *Ci-dit, Colloque international* 11-13 juin 2009, Nice.