

Des mots pour le dire : le discours des partis politiques suisses

Jacques Savoy

Institut d'informatique, Université de Neuchâtel

L'année 2007 a été marquée par les élections fédérales et la présidentielle française. Ces événements ont été accompagnés d'un flot de discours rédigés afin de rallier les électeurs ou d'indiquer les choix politiques que les formations entendent suivre. Comme l'électeur, nous avons comparé ces discours politiques mais, dans notre cas, à l'aide de l'informatique et de la statistique. Ces outils nous permettent d'ouvrir de nouvelles perspectives dans l'analyse des différences de vocabulaire usités par les quatre grands partis suisses ainsi que les deux principaux candidats à l'Elysée. En effet, le choix des mots et des expressions n'est pas le simple fruit du hasard et chaque intervenant prend un soin particulier à rédiger son programme. De plus, comme ces discours politiques ont été rédigés durant la même période, utilisant la même langue et désirant atteindre des objectifs similaires, leur comparaison directe s'en trouve faciliter.

Dans ce but, nous avons constitué un corpus de discours politiques suisses en téléchargeant les documents disponibles sur les sites web des partis (17 et 18 décembre 2007). Ces textes présentent la plate forme électorale (élection fédérale du 21 octobre 2007) ou les positions du parti dans cette perspective. Tous les documents récupérés sont rédigés en langue française et disponible en format pdf. Afin de limiter quelque peu nos investigations, nous avons retenu les quatre grands partis présents dans le gouvernement fédéral, à savoir, en partant de la droite, l'UDC (union démocratique du centre), le PRD (parti radical démocratique), le PDC (parti démocrate-chrétien) et le PS (parti socialiste). Quelques statistiques concernant ce corpus sont reprises dans la table 1 avec, comme élément de comparaison un corpus de dépêches d'agence de l'ATS (agence télégraphique suisse) datant des années 1994 et 1995.

	ATS	PS	PDC	PRD	UDC
Taille en octet	145 321 359	238 414	339 307	181 629	613 256
Nombre de mots	22 414 143	36 097	50 358	26 687	90 578
Nombre de vocables	139 157	4 167	5 238	3 280	7 191
<i>hapax</i>	50 912	1 969	2 346	1 501	3 079
<i>hapax</i> en %	36,59 %	47,25 %	44,79 %	45,70 %	42,82 %
Documents		7	22	9	8
Site Web	www.ats.ch	www.sp-ps.ch	www.pdc.ch	www.prd.ch	www.svp.ch

Table 1 : Quelques statistiques sur notre corpus

Notre étude s'appuie sur les mots présents ou absents de ces discours et en recourant à l'informatique pour en faire les traitements nécessaires. Quelques remarques préliminaires s'imposent. Si pour l'être humain la segmentation d'un texte en mots ne soulève pas de difficulté

apparente, il n'en est pas de même pour l'ordinateur. Nous avons considéré comme mot, toute séquence de lettres et ou de chiffres. Cette définition ne semble pas soulever d'objection. Cependant, quelques imperfections demeurent. Ainsi la forme « chemin de fer » sera analysée comme trois mots (« chemin », « de », « fer ») et les termes « ne ... pas » ou « parce que » mériteraient d'être comptés sous une entrée unique. D'autre part, les formes « l'école » ou « aujourd'hui » seront vues comme composée de deux mots (soit « l » et « école » et « aujourd » et « hui »). Notre système ne fera pas de distinction entre majuscule et minuscule et les formes « Suisse » ou « suisse » seront considérées comme identiques. Certes, les formes « poste » et « Poste » devraient être vues comme des termes différents dans la phrase « le poste ouvert à la Poste » car ils recouvrent clairement deux entités distinctes. Toutefois, si un mot s'écrit exclusivement avec des majuscules, nous avons conservé cette forme en l'état. En effet, elle correspond souvent à des acronymes (UE, PRD, ONU, AI). Nous n'avons pas effectuée une analyse morphologique poussée afin de déterminer pour chaque mot son entrée dans le dictionnaire (lemmatisation). Dans notre cas, les formes « veux », « voulons » ne seront pas regroupées sous le même vocable « vouloir ». Nous avons toutefois appliqué un enracineur léger (Savoy 2002) supprimant la marque du pluriel (le « -s » final ou la transformation de la séquence finale « -aux » en « -al »). En règle générale, nous pensons que les calculs effectués divergent quelque peu par rapport à une lemmatisation complète mais les conclusions que nous en tirons devraient demeurer identiques, très similaires pour le moins.

L'analyse lexicométrique que nous allons appliquer relève l'identité de l'auteur (par exemple pour l'attribution d'une œuvre controversée (Labbé, 2007)), le courant de pensée d'un groupe ou la comparaison avec d'autres auteurs (Labbé & Monière, 2003). On reconnaîtra tout de même que le choix d'un vocabulaire peut également être sujet à des variations dues aux circonstances (le contexte, l'auditoire, intervention spontanée ou discours lue) ainsi qu'au type de communication choisi (programme général ou discours traitant une question particulière ou technique). Dans le contexte présent, ces diverses variations sont relativement neutralisées dans notre corpus. En effet, les textes proviennent de la même période, sont rédigés dans la même perspective et couvrent des objectifs très similaires.

Le vocabulaire des partis politiques suisses

Comme première analyse, nous pourrions comparer la taille du vocabulaire utilisé par les quatre intervenants. Dans cette perspective, nous pouvons estimer qu'un lexique étendu correspond à un parti ayant de grandes ambitions, désirant couvrir tous les domaines (Labbé & Monière, 2003). Mais en présence d'un vocabulaire plus restreint, nous pourrions avancer l'hypothèse que le parti a opté pour la sobriété, pour un parler simple et direct. Cependant, cette analyse doit s'effectuer sur un ensemble de documents possédant la même taille ou de longueur très similaire car un corpus possédant un volume plus important serait favorisé au regard de la taille du vocabulaire. Comme l'indique la table 1, les quatre grands partis présentent des volumes distincts. Comme le PRD propose le corpus le moins long (26 687 mots), nous avons réduit les trois autres corpus à cette dimension pour cette première analyse.

Si l'on compare, en prenant le même nombre de mots (soit 26 687), le nombre de formes différentes (vocables) utilisés par les quatre grands partis suisses, on retrouve le vocabulaire le plus étendue auprès de l'UDC avec 3 901 formes, suivi par le PDC avec 3 689 vocables, puis les 3 382 formes pour le PS et, finalement, 3 280 pour le PRD. A titre de comparaison, un corpus de même taille extrait des dépêches d'agence de l'ATS présente 4 359 vocables. La différence s'élève à environ 16 % entre les deux extrêmes (UDC et PRD), situé tous les deux à la droite du

parlement. Le parler simple et direct serait l'apanage du PRD tandis que les grandes ambitions et la couverture la plus large seraient plutôt du côté de l'UDC. D'un autre côté, si la rareté des mots serait un indice de la richesse lexicale avec des expressions savantes n'apparaissant qu'une seule fois, l'UDC remporte de nouveau le premier rang avec 1 960 vocables apparaissant qu'une seule fois (*hapax*) contre 1 800 pour le PDC, 1 650 au PS et 1 501 au PRD.

	PS	PDC	PRD	UDC	ATS
mots	26 687	26 687	26 687	26 687	26 687
vocable	3 382	3 689	3 280	3 901	4 359
hapax	1 650	1 800	1 501	1 960	2 094

Table 2 : Richesse lexicale en nombre de vocables (forme distincte) et nombre de vocables apparaissant qu'une seule fois (*hapax*)

Si l'on classe, pour chaque discours, les divers vocables selon leur fréquence d'apparition, nous observons l'usage fréquent des mêmes formes pour les quatre partis. Cependant, un regard plus attentif révèle que ces formes très fréquentes correspondent à des mots-outils ou mots grammaticaux tels que déterminants (la, le, des, ce), prépositions (à, de, par, dans, pour), conjonctions (et, car, ou, que) ainsi que quelques formes verbales (est, ont, été). Après élimination de cet ensemble de mots peu ou pas porteurs de sens, nous voyons mieux émerger les thèmes privilégiés propres à chaque parti comme l'illustre la table 3.

PS	PDC	PRD	UDC
nous 19	nous 10	suisse 22	suisse 13
politique 21	suisse 20	doit 23	politique 26
doit 24	PDC 33	politique 25	AI 31
suisse 28	politique 36	nous 26	droit 35
culturelle 35	doit 37	sécurité 27	UDC 38
culture 37	formation 40	doivent 34	état 41
sociale 42	jeune 42	armée 35	étranger 42
droit 43	enfant 43	PRD 37	doit 43
PS 44	doivent 44	cadre 41	fédéral 44
travail 46	pays 48	domaine 33	naturalisation 48

Table 3 : Les dix vocables les plus fréquents (et leur rang) dans les quatre partis

Dans la table 3, nous avons repris pour chaque parti les dix vocables les plus fréquents et possédant une sémantique propre. A côté de chacune des formes, nous avons également indiqué son rang parmi les vocables les plus fréquents. Nous pouvons constater que certains mots apparaissent fréquemment dans les quatre discours comme « politique », « doit » ou « suisse ». Si l'on analyse ce dernier vocable, on constate que son rang diverge quelque peu entre les partis. Pour le PDC ou le PRD, ce terme possède le rang 20 ou 22, tandis que ce vocable semble moins utilisé au PS (rang 28). Par contre le vocable « suisse » dispose d'un rang élevé à l'UDC (rang

13), indiquant plus clairement la volonté de ce parti d'axer son discours sur la Suisse, face à l'étranger (rang 42) ou la naturalisation (rang 48).

Les programmes des quatre partis gouvernementaux font fréquemment référence à eux-mêmes et, dans une moindre mesure, aux autres partis. Selon notre décompte, chaque parti fait essentiellement référence à lui-même. Par exemple, dans les documents décrivant le programme du PRD, le sigle « PRD » apparaît en 37^e position (avec une fréquence de 85), tandis que dans le site du PS, son sigle apparaît en 44^e position avec une fréquence d'occurrence de 101 (l'acronyme PSS pour « parti socialiste suisse » n'apparaît lui que deux fois). La règle se confirme avec les deux autres partis, à savoir le PDC (33^e position, fréquence de 194) ou l'UDC (38^e position, fréquence de 286). Dans les documents analysés, les partis ne feront que peu référence aux autres formations. Ainsi pour le PS ou le PDC, le sigle des autres partis n'apparaît pas de manière notable avec des fréquences d'occurrence entre 12 et 3. Remarquons toutefois deux exceptions. Les documents du parti PRD n'utilisent jamais le sigle d'un autre parti tandis que l'abréviation « PS » apparaît de manière fréquente dans les programmes de l'UDC (soit en 106^e position avec 100 occurrences).

La politique c'est aussi l'affectation des ressources et donc l'usage de mots comme « franc », « million » ou « milliard ». Sur cette thématique, l'UDC prend la tête en ayant le mot « franc » qui apparaît en 50^e position (206 occurrences) puis le mot « milliard » en 59^e rang (fréquence absolue de 169). Le terme « million » reste un peu en deçà avec un rang de 161 (fréquence de 67). Le discours financier reste une caractéristique de l'UDC, les autres partis présentent des rangs moins élevés pour ces divers vocables. Ainsi, en position 140 du parti PDC, on retrouve le mot « franc » (fréquence de 43) puis le mot « million » (rang : 379, 18 occurrences) et le terme « milliard » (rang : 451, fréquence 455). Au PS, le mot « million » apparaît au 214^e rang (23 occurrences) suivi par le terme « milliard » (267^e position, fréquence 19) et le mot « franc » (rang : 229, fréquence 22). Au PRD, on se limite aux mots « franc » et « million », les deux formes ayant le même rang (731) pour une fréquence d'apparition de 5. Étonnamment, le terme « milliard » n'est pas utilisé dans les textes du programme du vieux parti.

La présence de mots de langue étrangère donne également quelques indications sur le parti qui en a recours. Ainsi du côté du PRD, on rencontre facilement des mots d'origine anglaise avec « tax » (position 215, 17 occurrences), « swiss » (position 232) ou « easy » (position 233). Ce trio n'est pas étonnant car il correspond à un thème cher à ce parti avec l'expression « Easy Swiss Tax ». L'UDC ouvre également son vocabulaire à des expressions étrangères mais dans ce cas précis, il s'agit de nommer les sources utilisées comme « Anzeiger » (position 729, fréquence de 15) dans le nom « Tages Anzeiger » (voire « Weltwoche » (position 1308, pour 8 occurrences)). Notons toutefois que le mot « Konzept » (rang 1081, fréquence absolue de 10) note bien un ancrage vers la langue allemande en dehors des références journalistiques. Le terme « swiss » apparaît en 730^e position au PS (avec 6 occurrences). Finalement, l'indication des sources semble une pratique bien ancrée dans les documents de l'UDC. Ainsi le terme « OFS » (office fédérale de la statistique) apparaît uniquement dans les textes de l'UDC, en position 691 (16 occurrences), tandis que la forme allemande (« bfs » pour Bundesamt für Statistik) n'apparaît qu'une seule fois.

Si l'on étudie les noms de villes ou de cantons, c'est le mot « Zürich » qui revient le plus fréquemment dans les partis PDC, PS et UDC (fréquence totale dans les trois programmes de 47). Les textes du PRD ne laissent pas transparaître des noms de villes ou de cantons. Suivent les mots de « Bâle » (24 apparition dans les programmes), « Berne » (21 occurrences) et « Genève » (14 occurrences). Si on élargit notre champ d'investigation à d'autres dénominations géographiques,

le mot « Europe » s'avère fréquent dans les programmes. Ainsi on le retrouve dans le parti PDC (rang 220, 30 occurrences) suivi par l'UDC (rang 401, fréquence de 28), le PRD (rang 384, fréquence de 10) et le PS (rang 45645, fréquence de 8).

Le thème des assurances sociales se profile dans la campagne électorale avec les mots « AVS » (assurance-vieillesse et survivants) (PDC : position 603, fréquence de 11, PRD : position 537, 7 occurrences ; PS : position 564, fréquence de 8 ; UDC : position 495, fréquence de 23) et « AI » (PDC : position 603, fréquence de 11 ; PRD : position 884, fréquence de 4 ; UDC : position 31, fréquence de 329).

Différences significatives entre partis

Si l'on désire connaître le vocabulaire plus spécifique à un parti, il faut le comparer à une norme, à un corpus politique suisse typique. Avec un tel point de référence, nous pouvons observer quel vocable apparaît de manière significativement plus fréquente dans l'un ou l'autre des discours ou, inversement, ceux dont l'occurrence s'avère significativement moins forte. Pour atteindre cet objectif, Muller (1992, chapitre 12) présente une méthodologie que nous avons suivie. Nous avons créé un corpus politique suisse typique en regroupant les documents rédigés par les quatre grands partis. Sur cette base, nous avons pu déterminer les formes sur-employées pour chaque formation politique (table 4) ainsi que celles qui sont sous-représentées dans leur discours respectif (table 5).

PS	PDC	PRD	UDC
état	nous	PRD	AI
culturelle	PDC	radical	UDC
II	demandons	armée	naturalisation
culture	énergie	défense	neutralité
artiste	internet	sécurité	gauche
PS	enfant	mission	rente
encouragement	notre	easy	nationalité
art	énergétique	militaire	état
autogestion	thème	imposition	milliard
CO2	électronique	cadre	franc

Table 4 : Les dix vocables surreprésentés de manière significative dans les quatre discours

Les vocables caractéristiques des divers partis indiqués dans la table 4 donnent plus clairement les grands thèmes privilégiés par chaque formation. Pour l'UDC, les thèmes récurrents sont les assurances sociales (« AI » assurance-invalidité et « rente »), la politique de naturalisation, la neutralité de la Suisse, l'affectation des ressources financières (« milliard », « franc ») mais également le souci de se distinguer de la « gauche ». Pour le PRD, les sujets de sécurité (« armée », « défense », « sécurité », « mission », « militaire ») forme une thématique centrale ainsi que les questions d'imposition fiscale (« easy » dans « easy swiss tax » ou « imposition»). Le terme « cadre » s'avère trop fortement ambigu et apparaît dans des contextes trop différents (« dans le cadre de la révision », « les conditions cadre », etc.) pour que nous puissions en tirer quelques

indications pertinentes. Le PDC centre son discours sur la famille (« enfant ») mais de manière un peu surprenante sur l'énergie et l'environnement (« énergie », « énergétique ») d'une part et, d'autre part, sur la technologie (« internet », « électronique »). Le PS semble se caractériser par sa politique culturelle (« culture », « artiste », « art ») mais également par une préoccupation écologique (taxe sur le « CO2 ») à côté d'un thème plus traditionnel (« autogestion »).

De manière duale, les vocables pas ou peu usités dans chaque parti permettent de compléter ces conclusions. Ainsi, on constate que les sigles des autres partis ne sont que très peu fréquents dans le discours d'un parti donné. On ne compare pas ou on n'attaque pas directement ses adversaires, à l'exception de l'UDC avec ses vocables « PS » et « gauche ». Les termes « naturalisation », « étranger » ou « milliard » sont visiblement des termes attachés à l'UDC. Etonnamment, le terme « neutralité » est sous-employé par le parti PRD dont le thème central concernait la sécurité et l'armée. Les mots « école » et « formation » sont également peu fréquents respectivement dans les partis PS et UDC. Enfin, le PS n'utilise que peu le terme « gauche ».

PS	PDC	PRD	UDC
école	milliard	problème	cadre
neutralité	étranger	énergie	II
jeune	état	milliard	état
PDC	PS	année	formation
enfant	rente	neutralité	demandons
gauche	armée	culturelle	économique
naturalisation	culture	naturalisation	voulons
suisse	UDC	culture	notre
UDC	culturelle	franc	PDC
état	AI	AI	nous

Table 5 : Les dix vocables sous-employés dans les discours des quatre partis

Distance intertextuel entre les discours

Ces premiers éléments nous permettent, sur la base du vocabulaire et de ses emplois spécifiques, de distinguer le discours des quatre grands partis suisses. Comme analyse complémentaire, nous pouvons calculer une distance entre les corpus pour connaître, par une mesure globale, les distances entre les divers partis. Pour atteindre cet objectif, nous devons proposer une mesure de distance entre deux textes. Par exemple, on peut compter le nombre de mots communs à deux documents divisé par la taille du vocabulaire. A cette pratique, on peut relever deux objections. Premièrement, les textes ne sont pas tous de la même ou d'une longueur similaire. Deuxièmement, la fréquence d'emploi d'un mot n'est pas prise en compte. Ainsi si un auteur utilise une seule fois le mot « abatement » et que celui-ci est très fréquent chez le second écrivain, la mesure n'en tiendrait pas compte.

Afin de résoudre ces objections, une mesure de distance entre textes (Brunet, 2003) a été proposée par Labbé & Labbé (2003) ou Labbé (2007), mesure que nous avons reprise dans nos calculs.

Avec cette technique, les valeurs possibles vont de 0,0 (les deux documents possèdent le même vocabulaire) à 1,0 (pas de mots communs entre les deux vocabulaires).

La table 6 indique les valeurs de distance que nous avons obtenu en comparant deux à deux les discours tenus par les quatre grands partis suisses, valeurs que nous avons complétées avec notre corpus de dépêche d'agence (ATS). Clairement, ce dernier corpus possède une distance plus grande avec les quatre discours politiques, indiquant une démarcation claire entre deux types de vocabulaire. Parmi les quatre discours politiques, les distances restent étonnamment proches, se situant dans des plages de valeurs de 0,29 à 0,31. Nous ne constatons pas une distance plus faible entre les partis de droite (distance (PRD, UDC) = 0,301) comparée aux extrêmes (distance (PS, UDC) = 0,291).

	ATS	PS	PDC	PRD	UDC
ATS	0	0,46776	0,45869	0,47012	0,45460
PS		0	0,30891	0,28460	0,29149
PDC			0	0,29602	0,30577
PRD				0	0,30097

Table 6 : Distance intertextuelle entre les discours politiques suisses

Comparaison Suisse - France

Comme l'année 2007 a également connu l'élection présidentielle française, nous pouvons comparer les discours électoraux tenus dans les deux pays. Comme l'indique la table 7, l'ordinateur révèle des différences s'avérant toutefois peu surprenantes. Ainsi, on privilégie les formes « suisse », « fédéral », « confédération », « franc » ou « canton » dans un pays et « français » « france » ou « république » dans l'autre. D'autres vocables relèvent les caractéristiques intrinsèques du discours électoral de S. Royal ou de N. Sarkozy. On y retrouve les formes verbales « crois », « veux », « suis » ou « dire » ou la conjonction « parce que » indiquant un besoin explicatif indéniable du côté français. En Suisse, on retrouve en surabondance les formes verbales « doit » ou « doivent » soulignant les obligations ou attentes (« l'Etat doit ») ainsi que les formes « économie », « assurance » ou « culturelle ». La table 8 permet de compléter cette analyse en donnant les vocables sur employés dans les deux pays.

Suisse	France	S. Royal	N. Sarkozy
suisse 1	je 1	je 1	je 1
nous 2	nous 4	nous 2	france 4
politique 4	france 5	france 5	nous 5
doit 7	veux 7	tous 8	veux 6
droit 9	tous 9	veux 10	si 7
pays 11	si 10	politique 17	français 10
fédéral 14	politique 13	jeune 18	politique 12
étranger 17	français 16	si 23	tous 13
formation 20	faire 18	emploi 26	faire 14

état 21	travail 29	pays 28	république 25
---------	------------	---------	---------------

Table 7 : Les dix vocables les plus fréquents (et leur rang) en Suisse et en France

Parfois les petits mots font toute la différence et dans ce cas nous rencontrons un emploi marqué du « je » (ainsi que du vocable relié « j ») dans les discours électoraux français par rapport à ceux de la Suisse. Pour Labbé & Monière (2008), le vocable « je » indique bien l'importance attachée à une personne, au chef du parti dans l'Hexagone. Plus étonnant, la fréquence d'occurrence du pronom « je » s'avère statistiquement plus élevée pendant le deuxième tour de la campagne présidentielle que lors du premier tour. Le passage au second tour note bien un changement au niveau lexical et cela pour les deux candidats. Pour un chef de parti, une campagne ne forme pas un continuum lexical stable, mais des ruptures peuvent apparaître. En fin de course, il faut serrer les rangs autour du « moi », du chef qui insistera sur le « je veux ».

Suisse	France
suisse	dire
fédéral	république
confédération	ai
neutralité	parce
naturalisation	français
PS	pas
canton	veux
franc	vous
nationalité	france
révision	je

Table 8 : Les dix vocables surreprésentés de manière significative dans les deux pays

Dans un deuxième temps, nous avons mesuré la distance entre les discours des quatre partis gouvernementaux suisses avec le discours des deux derniers candidats à l'Elysée (voir table 9). Dans ce but nous avons éliminé les vocables directement liés au pays (« suisses », « france », « français » « françaises »). Comparé aux distances entre partis suisses (voir table 6), les valeurs sont plus fortes entre les discours des deux pays (distance (Royal, Sarkozy) = 0,284). Si l'on compare les discours électoraux de S. Royal et de N. Sarkozy, on constate que le discours électoral de Ségolène se rapproche plus de trois des quatre formations suisses tandis que celui du président élu tend à se rapprocher plus de celui du PS. Comme seconde surprise, on constate que le programme de l'UDC se rapproche plus de celui de la gauche française.

	PS	PDC	PRD	UDC
N. Sarkozy	0,31297	0,40233	0,43989	0,39128
S. Royal	0,38754	0,34473	0,39843	0,36946

Table 9 : Distance intertextuelle entre les discours politiques suisses et français

Conclusion

Notre analyse du vocabulaire usité par les formations politiques suisses révèle bien les thèmes porteurs de ceux-ci mais également des nouvelles intentions (la technologie au PDC, la défense chez les radicaux, la culture au PS). Seul l'UDC indique clairement et fréquemment le sigle d'un autre parti (« PS » ou « gauche »). L'affectation des moyens financiers et l'intérêt à spécifier une source sont également deux caractéristiques de l'UDC. Par rapport à la campagne présidentielle française, les discours suisses sont plus similaires entre eux qu'avec celui d'un des candidats à l'Elysée. Ces derniers préférant les formes « je », « dire » et « parce que » dénotant l'importance du chef et un souci explicatif indéniable du côté français.

Références

- Yang C.C., Peng C.K., Yien H.W. & Goldberger A.L. : Information categorization approach to literary authorship disputes. *Physica: Statistical Mechanics and its Applications*, 329(3-4), 473-483, 2003.
- Brunet E. : Peut-on mesurer la distance entre deux textes ? *Corpus*, 2, 2003
- Labbé C. & Labbé D. : La distance intertextuelle. *Corpus*, 2, 2003
- Labbé C. & Monière D. : Le discours gouvernemental, Canada, Québec, France (1945-2000). Champion, Paris, 2003.
- Labbé D. : Experiments on authorship attribution by intertextual distance in English. *Journal of Quantitative Linguistics*. 14(1), 33-80, 2007.
- Labbé D., & Monière D. : Je est-il un autre ? Actes JADT 2008 (Journées internationales d'Analyse statistique des Données Textuelles), 647-656, mars 2008.
- Muller C. : Principes et méthodes de statistique lexicale. Champion, Paris, 1992.
- Savoy J. : Recherche d'informations dans des corpus en langue française : Utilisation du référentiel Amarylis. *TSI, Technique et Science Informatiques*, 21(3), 345-373, 2002.