

Projet TextCOOP

Typologies textuelles pour le traitement automatique des
documents électroniques : problèmes et enjeux

[Rapport d'analyse]

Maria ZIMINA-POIROT

(Version du janvier 2007)

LIPN – Laboratoire d'Informatique de l'université Paris-Nord (UMR7030)

Références :

- [1] ADAM, J.-M. (1992). *Les textes : types et prototypes - Récit, description, argumentation, explication et dialogue*. Paris : Nathan.
- [2] ADAM, J.-M. (1999). *Linguistique textuelle : Des genres de discours aux textes*. Paris : Nathan.
- [3] ARISTOTE (1932). *Rhétorique I*, texte établi et traduit par M. Dufour. Paris : Les Belles Lettres.
- [4] ARGAMON, SH., KARLGREN, J., SHANAHAN, J. G. (Eds.) (2005). « Stylistic Analysis Of Text For Information Access. » *Papers from the workshop held in conjunction with the 28th Annual International ACM Conference on Research and Development in Information Retrieval*, Salvador, Bahia, Brazil, August 13-19, 2005. Disponible sur : http://www.sics.se/jussi/Artiklar/2005_SIGIR_Salvador/Style2005/style2005.pdf
- [5] AUSSENAC-GILLES, N., CONDAMINES, A., SZULMAN, S. (2002). « Prise en compte de l'application dans la constitution de produits terminologiques. » In J. Le Maitre (Eds.), *Information, Interaction, Intelligence : Actes des 2e Assises Nationales du GDR I3*, Cépaduès Editions, p. 289-303. Disponible sur : <http://www.irit.fr/GDR-I3/fichiers/assises2002/papers/17-ConstitutionDeProduitsTerminologiques.pdf>
- [6] BAKHTINE, M. (1929/1977). *Marxisme et philosophie du langage*. Paris : Minuit.
- [7] BAKHTINE, M.(1984). *Esthétique de la création verbale*. Paris : Gallimard.
- [8] BEACCO, J-C. (1992) « Les genres textuels dans l'analyse du discours. » *Langages* n. 105, « Ethnolinguistique de l'écrit ».
- [9] BEAUDOUIN, V., FLEURY, S., HABERT, B., ILLOUZ, G., LICOPPE, C., PASQUIER, M. (2001). « TypWeb : décrire la Toile pour mieux comprendre les parcours. » In *Actes du Colloque International sur les Usages et les Services des Télécommunications, e-Usages CIUST'01*, Paris, 12-14 juin 2001. Disponible sur : <http://www.cavi.univ-paris3.fr/ilpga/ilpga/sfleury/Articles/typweb.pdf>
- [10] BEAUVISAGE, Th. (2004). « Sémantique des parcours des utilisateurs sur le web. » *Thèse de Doctorat en Sciences du langage*. Université de Paris X – Nanterre. Disponible sur : http://www.revue-texto.net/Inedits/Beauvisage/Beauvisage_Parcours.html

- [11] BELLOT, P., EL-BÈZE, M. (2001). « Classification et segmentation de textes par arbres de décision. Application à la recherche documentaire. » *Technique et science informatiques* n. 1(20), p. 107-134. Disponible sur : http://www.lia.univ-avignon.fr/fich_art/374-BellotElBeze1584.pdf
- [12] BENAMARA, F., MOENS, M-F., SAINT-DIZIER, P. (Eds.) (2005). *Proceedings of IJCAI'05 Workshop « Knowledge and Reasoning for Answering Questions. »* Edinburgh, July 30, 2005. Disponible sur : <http://www.irit.fr/recherches/ILPL/kraq05V1.pdf>
- [13] BEN ROMDHANE, M. (2001). « Navigation dans un espace textuel, accès à l'information scientifique ». *Thèse en Sciences de l'Information et de la Communication*. Université Jean Moulin Lyon 3. Disponible sur : http://www.recodoc.univ-lyon1.fr/these_MBR.pdf
- [14] BEUST, P., ROY, Th. (2006). « Prendre en compte la dimension globale d'un corpus dans la contextualisation du sens : expérimentations en informatique linguistique. » *GLOTTOPOL* n. 8 « Traitements automatisés des corpus spécialisés : contextes et sens. » juillet 2006. Disponible sur : http://www.univ-rouen.fr/dyalang/glottopol/telecharger/numero_8/gpl8_05beust_roy.pdf
- [15] BIBER, D. (1988). *Variations across speech and writing*. Cambridge : Cambridge University Press.
- [16] BIBER, D. (1995). *Dimensions of register variation*. Cambridge : Cambridge University Press.
- [17] BIBER, D. (2004). « Conversation text types: A multi-dimensional analysis. » In *Actes des 7es Journées internationales d'Analyse statistique des Données Textuelles JADT'04*, Louvain La Neuve, 10-12 mars, 2004, p. 15-34. Disponible sur : http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2004/pdf/JADT_000.pdf
- [18] BOMMIER-PINCEMIN, B. (1999). « Diffusion ciblée automatique d'informations : conception et mise en oeuvre d'une linguistique textuelle pour la caractérisation des destinataires et des documents. » *Thèse de Doctorat en Sciences du Langage*. Université Paris 4. Disponible sur : http://www.revue-texto.net/Inedits/Pincemin/Pincemin_these.html
- [19] BRANCA-ROSOFF, S. (1999). « Types, modes et genres entre langue et discours. » *Langage et société* n. 87, p. 5-24. Disponible sur : <http://www.cavi.univ-paris3.fr/llpga/ed/dr/drsb/sb-pdf/intro-Branca-LS87.pdf>

- [20] BRONCKART, J.-P. (1996). « Genres de textes, types de discours et opérations psycholinguistiques. » *ENJEUX* n. 37/38, p. 31-47.
- [21] BRUNET, É. (1988). « Une mesure de la distance intertextuelle : la connexion lexicale. » *Revue Informatique et Statistique dans les Sciences humaines RISSH* n. 24(1-4), p. 81-116.
- [22] BRUNET, É. (2000). « Qui lemmatise, dilemme attise. » *Revue électronique Lexicometrica 2*. Disponible sur : <http://www.cavi.univ-paris3.fr/lexicometrica/article/numero2/brunet2000.PDF>
- [23] BRUNET, É. (2003). « Peut-on mesurer la distance entre deux textes ? » *CORPUS* n. 2 « La distance intertextuelle. » décembre 2003.
Disponible sur : <http://corpus.revues.org/document30.html>
- [24] BRUNET, É. (2004). « Où l'on mesure la distance entre les distances. » *Texte ! Dits et Inédits*. mars 2004.
Disponible sur : http://www.revue-texto.net/Inedits/Brunet/Brunet_Distance.html
- [25] BRUNET, E. (2006). « Navigation dans les rafales. » In *Actes des 8es Journées internationales d'Analyse statistique des Données Textuelles JADT'06*, Besançon, 19-21 avril 2006, p. 15-29. Disponible sur : http://www.cavi.univ-paris3.fr/lexicometrica/jadt/JADT2006-PLENIERE/JADT2006_EB.pdf
- [26] CEHEUX, G. R. (nom collectif) (2002). « Stratégie pour l'interprétation de documents. » In *Actes des deuxièmes assises nationales du GdR I3 (Information-Interaction-Intelligence)*, p. 275-288. Disponible sur : <http://www.irit.fr/GDR-I3/fichiers/assises2002/papers/16-InterpretationDeDocuments.pdf>
- [27] CHARAUDEAU, P. (1997). « Les conditions d'une typologie des genres télévisuels d'information. » *Réseaux* n. 81, CNET. Disponible sur : <http://www.enssib.fr/autres-sites/reseaux-cnet/81/05-chara.pdf>
- [28] CHARAUDEAU, P., MAINGUENEAU, D. (2002). *Dictionnaire d'analyse du discours*. Paris : Seuil.
- [29] CHAROLLES, M. (1988). « Les plans d'organisation textuelle, périodes, chaînes, portées et séquences. » *Pratiques* n. 57, Metz, p. 3-14.
- [30] CLAVIER, V. (2006). « Le genre comme point d'accès au document : analyse comparée de textes scientifiques en mécanique et linguistique. » In *Contributions à la Journée ATALA*

« *Typologies de textes pour le traitement automatique* », Paris, 9 décembre 2006.
Disponible sur : http://www.atala.org/article.php3?id_article=312

[31] COPECK, T., BARKER, K., DELISLE, S., SZPAKOWICZ, S. (2000). « Automating the Measurement of Linguistic Features to Help Classify Texts as Technical. » In *Actes de la 7ème conférence annuelle sur le Traitement Automatique des Langues Naturelles TALN'2000*, Lausanne, 16-18 octobre, 2000, p. 101-110. Disponible sur :
<http://www.cs.utexas.edu/~kbarker/papers/taln00-tt.pdf>

[32] COSTA, R. (2005). « Texte, terme, contexte. » In Actes des 7es Journées scientifiques du Réseau de chercheurs 'Lexicologie, Terminologie, Traduction', Bruxelles, 8-10 septembre 2005, p. 79-87.

Disponible sur : <http://perso.univ-lyon2.fr/~thoiron/JS%20LTT%202005/pdf/Costa.pdf>

[33] CROWSTON, K., KWASNIK, B. (2004). « A Framework for Creating a Faceted Classification for Genres: Addressing Issues of Multidimensionality. » In *Proceedings of the 37th Hawaii International Conference on System Sciences HICSS'04*, Big Island, Hawaii, 9 p.
Disponible sur :

<http://csdl2.computer.org/comp/proceedings/hicss/2004/2056/04/205640100a.pdf>

[34] DEWDNEY, N., VANESS-DYKEMA, C., MACMILLAN, R. (2001). « The form is the substance: Classification of genres in text. » In *Proceedings of the ACL Workshop on Human Language Technology and Knowledge Management*, Toulouse, France, 6-7 Juillet 2001, p. 142–149. Disponible sur : <http://www.elsnet.org/km2001/dewdney.pdf>

[35] ERTZSCHEID, O. (2002). « L'hypertexte : haut lieu de l'intertexte. » *Revue des ressources* [revue électronique publié par Points de Fuite multimédia]. octobre 2002.
Disponible sur : http://www.larevuedesressources.org/article.php3?id_article=27

[36] FINN, A., KUSHMERICK, N. (2003). « Learning to classify documents according to genre. » In *Proceedings of IJCAI-03 Workshop on Computational Approaches to Style Analysis and Synthesis*, Acapulco, Mexico, August 10, 2003, p. 1506-1518. Disponible sur :
<http://www.aidanf.net/publications/finn03learninggenre.pdf>

[37] FLØTTUM, K., HOLM, H-V. (Eds.) (1999). « Polyfoni. » *TRIBUNE* n.9, Skriftserie for Romansk Institutt, Universitetet i Bergen. Disponible sur :
<http://www.hum.au.dk/romansk/polyfoni/>

[38] GENETTE, G. (1987). *Seuils*. Paris : Éditions du Seuil.

- [39] GREIMAS, A.-J. (1966). *Sémantique structurale*. Paris : Larousse.
- [40] HABERT, B., ILLOUZ, G., LAFON, P., FLEURY, S., FOLCH, H., HEIDEN, S., PREVOST, S. (2000). « Profilage de textes : cadre de travail et expérience. » In *Actes des 5es Journées internationales d'Analyse des Données Textuelles JADT'00*, Lausanne, 9-11 mars 2000.
Disponible sur : <http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2000/pdf/56/56.pdf>
- [41] HABERT, B., FABRE, C., ISSAC, F. (1998). *De l'écrit au numérique. Constituer, normaliser et exploiter les corpus électroniques*. Paris : Masson.
- [42] HABERT, B., NAZARENKO, A., SALEM, A. (1997). *Les linguistiques de corpus*. Paris : Armand Colin / Masson.
- [43] HABERT, B., SALEM, A. (1995). « L'utilisation de catégorisations multiples pour l'analyse quantitative de données textuelles. » *TAL* n. 36(1-2), p. 249-276.
- [44] HERNANDEZ, N. (2004). « Description et Détection Automatique de Structures de Textes. » *Thèse de Doctorat en Informatique*. Université Paris-Sud.
- [45] HOLZEM M., DIONISI, D., LABICHE, J., TRUPIN, E. (2005). « Le document dans son agir organisationnel : le modèle de l'organisation dans l'interaction usager système. » In *Actes du huitième colloque international sur le document électronique CIDE'05*, Beyrouth, Liban, 25-28 mai 2005, p. 133-151. Disponible sur :
http://archivesic.ccsd.cnrs.fr/docs/00/06/25/90/PDF/sic_00001381.pdf
- [46] ILLOUZ, G. (1999a). « Méta-Étiqueteur Adaptatif : Vers une utilisation pragmatique des ressources linguistiques. » In *Actes de l'Atelier thématique TALN'99 : Corpus et Traitement Automatique des Langues : pour une réflexion méthodologique*, Cargèse, 12-17 juillet 1999, p. 15-24. Disponible sur : <http://www.limsi.fr/Individu/gabrieli/CV/Publis/Articles/taln99.pdf>
- [47] ILLOUZ, G., HABERT, B., FLEURY, S., FOLCH, H., HEIDEN, S., LAFON, P. (1999b). « Maîtriser les déluges de données hétérogènes. » In *Actes de l'Atelier thématique TALN'99 : Corpus et Traitement Automatique des Langues : pour une réflexion méthodologique*, Cargèse, 12-17 juillet 1999, p. 37-46. Disponible sur :
<http://www.limsi.fr/Individu/gabrieli/CV/Publis/Articles/taln99-typweb.pdf>
- [48] ILLOUZ, G. (2000a). « Vers un apprentissage en TALN dépendant du type de Texte. » In *Actes de la 7ème conférence annuelle sur le Traitement Automatique des Langues*

Naturelles TALN'2000, Lausanne, 16-18 octobre 2000. Disponible sur :

<http://www.limsi.fr/Individu/gabrieli/CV/Publis/Articles/FinaleTALN2000illouz.pdf>

[49] ILLOUZ, G., HABERT, B., FLEURY, S., FOLCH, H., HEIDEN, S., LAFON, P., PREVOST, S. (2000b). « Typtex : Generic Features for Text Profiler. » In *Proceedings of 6th RIAO Conference on "Content-Based Multimedia Information Access"*, Paris, 12-14 avril 2000, p.1526-1540. Disponible sur :

<http://www.limsi.fr/Individu/gabrieli/CV/Publis/Articles/RIAO2000-TyPTex.ps.gz>

[50] JAILLET, S. (2004). « Catégorisation automatique de documents. » In *Actes de la douzième session des journées des doctorants DOCTISS'04*, ED « Information, Structures et Systèmes », Université Montpellier 2.

Disponible sur : <http://www.lirmm.fr/doctiss04/art/I02.pdf>

[51] JAILLET, S., TEISSEIRE, M., CHAUCHE, J., PRINCE, V. (2003). « Classification automatique de documents. Le coefficient des deux écarts. » In *Actes du XXIème Congrès INFORSID'2003*, Nancy, 3-6 juin 2003. p. 87-102. Disponible sur :

<http://134.214.81.35/articles/a447c1bguEFeLWdKY.pdf>

[52] JALAM, R. (2003). « Apprentissage automatique et catégorisation de textes multilingues. » *Thèse de Doctorat en Informatique*. Université Lumière Lyon 2. Disponible sur : http://www.agrocampus-rennes.fr/math/jalam/these/these_radwan.pdf

[53] JOACHIMS, T. (1998). « Text categorization with support vector machines: learning with many relevant features. » In *Proceedings of 10th European Conference on Machine Learning ECML-98*, p. 137–142. Disponible sur :

http://www.cs.cornell.edu/People/tj/publications/joachims_98a.ps.gz

[54] KARLGRÉN, J., CUTTING, D. (1994). « Recognizing Text Genres with Simple Metrics Using Discriminant Analysis. » In *Proceedings of the 15th International Conference on Computational Linguistics COLING'94*, p. 1071-1075.

Disponible sur : <http://eprints.sics.se/56/01/cmplglixcol.pdf>

[55] KASTBERG SJÖBLOM, M. (2006). « La variation typologique : analyse systématique d'un corpus québécois. » In *Contributions à la Journée ATALA « Typologies de textes pour le traitement automatique »*, Paris, 9 décembre 2006.

Disponible sur : http://www.atala.org/article.php3?id_article=312

[56] KESSLER, B., NUNBERG, G., SCHÜTZE, H. (1997). « Automatic Detection of Text Genre. » In *Proceedings of the Thirty-Fifth Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, Madrid, Spain, July 7-12 1997, p. 32-38.

Disponible sur : <http://acl.ldc.upenn.edu/P/P97/P97-1005.pdf>

[57] LAMALLE, C., SALEM, A. (2002). « Types généralisés et topographie textuelle dans l'analyse quantitative des corpus textuels. » In *Actes des 6es Journées internationales d'Analyse statistique des Données Textuelles JADT'2002*, Saint-Malo, 13-15 mars 2002, p. 403-411. Disponible sur : http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2002/PDF-2002/lamalle_salem.pdf

[58] LEBART, L., SALEM, A. (1994). *Statistique textuelle*. Paris : Dunod.

[59] LEBLANC J-M., MARTINEZ, W. (2005). « Positionnements énonciatifs dans les vœux présidentiels sous la cinquième République ». *CORPUS* n. 4 « Les corpus politiques : objet, méthode et contenu. » décembre 2005. Disponible sur :

<http://corpus.revues.org/document347.html>

[60] LEE, D. (2001). « Genres, Registers, Text Types, Domains, and Styles: Clarifying the Concepts and Navigating a Path through the BNC Jungle. » *Language Learning & Technology* n. 3(5), p. 37-72. Disponible sur : <http://llt.msu.edu/vol5num3/lee/>

[61] LIAO, C., ALPHA, Sh., DIXON, P. (2003). « Feature Preparation in Text Categorization. » In *Proceedings of the Second Australasian Data Mining Conference AusDM'03*, Canberra, Australia (in conjunction with the 2003 Congress on Evolutionary Computation CEC'2003), 8 December 2003. Disponible sur :

http://www.oracle.com/technology/products/text/pdf/feature_preparation.pdf

[62] LUNDQUIST, L., MINEL, J-L., COUTO, J. (2006). « NaviLire, Teaching French by Navigating in Texts. » In *Proceeding of Eleventh International Conference Information Processing and Management of Uncertainty in Knowledge-based Systems IMPU'2006*, Paris, July 2-7, 2006. p. 1093-1099. Disponible sur : http://halshs.archives-ouvertes.fr/docs/00/09/78/49/PDF/Lundquist_Minel_Couto_IPMU06.PDF

[63] LUONG, X. (dir.) (2003). *CORPUS* n. 2 « La distance intertextuelle. » décembre 2003. Disponible sur : <http://corpus.revues.org/sommaire52.html>

- [64] LUŠTREK, M. (2006). « Overview of automatic Genre Identification. » *Technical Report*. Jozef Stefan Institute, Slovenia. Disponible sur : http://dis.ijs.si/mitjal/documents/Overview_of_Automatic_Genre_Identification-TR-06.pdf
- [65] MALRIEU, D. (2001). « Stylistique et Statistique textuelle: À partir de l'article de C. Muller sur les "pronoms de dialogue" .» *Texte ! Dits et Inédits*. [Revue électronique en ligne]. Disponible sur : http://www.revue-texto.net/Inedits/Malrieu_Stylistique.pdf
- [66] MALRIEU D. (2004). « Linguistique de corpus, genres textuels, temps et personnes. *Langages* n. 153, p. 73-85. Disponible sur : http://infolang.u-paris10.fr/modyco/textes/malrieu/DM_Genres_temps_personnes_03.pdf
- [67] MALRIEU, D., RASTIER, F. (2001). « Genres et variations morphosyntaxiques. » *TAL* n. 2(42), p. 548-577. Disponible sur : http://www.revue-texto.net/Inedits/Malrieu_Rastier/Malrieu-Rastier_Genres1.html
- [68] MANGENOT, F. (1996). « Outils textuels pour l'apprentissage de l'écriture en L1 et en L2. » In *Actes du Xe colloque international FOCAL 'Pratiques discursives et acquisition des langues étrangères'*, Besançon, 19-21 septembre 1996, p. 515-525. Disponible sur : http://w3.u-grenoble3.fr/espace_pedagogique/focal.htm
- [69] MARANINCHI, L. (2001). « Etude de marqueurs linguistiques pour une interprétation sémantique du référent du pronom sujet 'nous' ». Mémoire *de maîtrise en Sciences du Langage*, mention Industrie de la Langue. Université de Paris III, Sorbonne Nouvelle. Disponible sur : <http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/sitespp/maitrise-2001/laetitia/memoireLM2001.pdf>
- [70] MARSHMAN, E. (2003). « Construction et gestion des corpus : Résumé et essai d'uniformisation du processus pour la terminologie. » Étude réalisée pour le groupe *ÉCLECTIK*. L'Observatoire de linguistique Sens-Texte, Université de Montréal. Disponible sur : <http://www.olst.umontreal.ca/pdf/terminotique/corpusnormes.pdf>
- [71] MARTINEZ, W., ZIMINA, M. (2002). « Utilisation de la méthode des cooccurrences pour l'alignement des mots de textes bilingues. » In *Actes des 6es Journées internationales d'Analyse statistique des Données Textuelles JADT'2002*, Saint-Malo, 13-15 mars 2002, p. 495-506. Disponible sur : http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2002/PDF-2002/martinez_zimina.pdf

- [72] MEYER ZU EISSEN, S., STEIN, B. (2004). « Genre Classification of Web Pages: User Study and Feasibility Analysis. » In Biundo S., Fruhwirth T., Palm G. (Eds.), *Advances in Artificial Intelligence*. Berlin : Springer, p. 256-269. Disponible sur : <http://www-ai.upb.de/aisearch/ki04-frame.pdf>
- [73] MOIRAND, S. (2003). « Quelles catégories descriptives pour la mise au jour des genres du discours ? » In *Contributions à la Journée d'études : les genres de l'oral*, Université Lumière – Lyon 2, le 18 avril 2003. Disponible sur : http://icar.univ-lyon2.fr/Equipe1/actes/Journee_Genre/Moirand_cat_genres.rtf
- [74] NAZARENKO, A. (2003). *Projet ExtraPloDocs : Rapport d'analyse de l'existant et des besoins*. Consortium LIPN, MIG, ISOFT. Rapport 2.1 et 2.2. Disponible sur : <http://www-lipn.univ-paris13.fr/~poibeau/Extra/D21.pdf>
- [75] OUERFELLI, T., LALLICH-BOIDIN, G. (2000). « Pratiques d'indexation dans les Bases Textuelles Structurées : Application aux Textes Techniques sous Format HTML. » In *Travaux du 28e congrès annuel de l'Association canadienne des sciences de l'information*. Disponible sur : <http://www.slis.ualberta.ca/cais2000/ouerfelli.htm>
- [76] PERY-WOODLEY, M-P. (1995). « Quels corpus pour quels traitements automatiques ? » *TAL* n. 36(1-2), p. 213-232. Disponible sur : <http://w3.univ-tlse2.fr/erss/membres/pery/articles/TALcorpus.pdf>
- [77] PERY-WOODLEY, M-P. (2000). « Une pragmatique à fleur de texte : approche en corpus de l'organisation textuelle. » *Mémoire d'HDR*, Université de Toulouse-LeMirail : ERSS, In *Carnets de grammaire* n. 8, juillet 2000. Disponible sur : <http://w3.univ-tlse2.fr/erss/membres/pery/articles/habilmppw.pdf>
- [78] PERY-WOODLEY, M-P. (2001). « Modes d'organisation et de signalisation dans des textes procéduraux. » *Langages* n. 141, p. 28-46. Disponible sur : <http://w3.univ-tlse2.fr/erss/membres/pery/articles/lang00.pdf>
- [79] PETITJEAN, A. (1989). « Les typologies textuelles. » *Pratiques* n. 62, p. 86-125.
- [80] PETITJEAN, A. (2005). « Pour une problématisation linguistique de la notion de genre textuel. » In *Actes du VIe Congrès des Romanistes Scandinaves*, Copenhague, 24-27 août 2005. Disponible sur : www.ruc.dk/isok/skriftserier/XVI-SRK-Pub/MOL/MOL07-Petitjean/
- [81] PIERARD, S, BESTGEN, Y. (2006). « A la pêche aux marqueurs linguistiques de la structure du discours. » In *Actes des 8es Journées internationales d'Analyse statistique des*

Données Textuelles JADT'06, Besançon, 19-21 avril 2006. Disponible sur :
<http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2006/PDF/II-067.pdf>

[82] POIBEAU, T., NAZARENKO, A. (1999). « L'extraction d'information, une nouvelle conception de la compréhension de texte ? » *TAL* n. 40(2).

[83] POU DAT C. (2006). « Etude contrastive de l'article scientifique de revue linguistique dans une perspective d'analyse des genres. » *Thèse de Doctorat en Sciences du langage*. Université d'Orléans.

Disponible sur : <http://www.revue-texto.net/Corpus/Publications/Poudat/>

[84] RASTIER, F. (1989). *Sens et textualité*. Paris : Hachette.

[85] RASTIER, F. (2006) « De l'origine du langage à l'émergence du milieu sémiotique. » *Marges linguistiques* n. 11. Disponible sur :

http://marg.lng1.free.fr/documents/13_ml112006_rastier_f/13_ml112006_rastier_f.pdf

[86] RATLIFF, E. (2006). « Me Translate Pretty One Day. » *WIRED* n. 14.12, December 2006.

Disponible sur :

http://www.wired.com/wired/archive/14.12/translate.html?pg=1&topic=translate&topic_set=

[87] RAUBER A., MULLER-KOGLER, A. (2001). « Integrating automatic genre analysis into digital libraries. » In *Proceedings of the First ACM-IEEE Joint Conference on Digital Libraries*. ACM Press, Roanoke, Virginia. Disponible sur :

http://www.ifs.tuwien.ac.at/ifs/research/pub_pdf/rau_jcdl01.pdf

[88] ROULET, E. (1991). « Une approche discursive de l'hétérogénéité discursive. » *Etudes de linguistique appliquée* n. 83, p. 117-130.

[89] SALEM, A. (1987). *Pratique des segments répétés : essai de statistique textuelle*. Paris : Klincksieck.

[90] SALEM, A. (2004). « Introduction à la résonance textuelle. » In *Actes des 7es Journées internationales d'Analyse statistique des Données Textuelles JADT'04*, Louvain La Neuve, 10-12 mars 2006, p. 986-992. Disponible sur :

http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2004/pdf/JADT_096.pdf

[91] SALEM, A. (2006). « Proximités segmentales. » In *Actes des 8es Journées internationales d'Analyse statistique des Données Textuelles JADT'06*, Besançon, 19-21 avril

2006, p. 986-992.

Disponible sur : <http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2006/PDF/II-075.pdf>

[92] SANTINI, M. (2003). « Identifying Genres on the Web: PhD Thesis Outline. » *ITRI report series: ITRI-03-06*, Brighton University, UK.

Disponible sur : <ftp://ftp.itri.bton.ac.uk/reports/ITRI-03-06.pdf>

[93] SANTINI, M. (2004). « State-of-the-art on Automatic Genre Identification. » *ITRI report series: ITRI-04-03*, Brighton University, UK.

Disponible sur : <ftp://ftp.itri.bton.ac.uk/reports/ITRI-04-03.pdf>

[94] SANTINI, M. (2005). « Linguistic Facets for Genre and Text Type Identification: A Description of Linguistically-Motivated Features. » *ITRI report series: ITRI-05-02*, Brighton University, UK. Disponible sur : <ftp://ftp.itri.bton.ac.uk/reports/ITRI-05-02.pdf>

[95] SANTINI, M. (2006). « Some Issues in Automatic Genre Classification of Web Pages. » In *Actes des 8es Journées internationales d'Analyse statistique des Données Textuelles JADT'06*, Besançon, 19-21 avril 2006, p. 865-876. Disponible sur : <http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2006/PDF/II-077.pdf>

[96] SCHNEUWLY, B. (1994). « Genres et types de discours: considérations psychologiques et ontogénétiques. » In Y. Reuter (Ed.), *Les interactions lecture-écriture*. Bern : Lang, p. 155-174.

[97] SEBASTIANI, F. (2002). « Machine learning in automated text categorization. » *ACM Computing Surveys* n. 34(1), p. 1-47. Disponible sur : http://br.endernet.org/~akrowne/elaine/dlib/papers/sebastiani/sebastiani_classification_survey.pdf

[98] SEKINES, S. (1997). « The domain dependence of parsing. » In *Proceedings of the fifth Conference on Applied Natural Language Processing*, Washington, p. 96–102. Disponible sur : <http://acl.ldc.upenn.edu/A/A97/A97-1015.pdf>

[99] SERANT, D., THOIRON, P. (1992). « Topographie des formes répétées. » *Lingua* n.3(87), p. 333-343.

[100] SILVA, R. (2005). « Morphologie de spécialité : regard(s) sur le(s) contexte(s). » In *Actes des 7es Journées scientifiques du Réseau de chercheurs 'Lexicologie, Terminologie, Traduction'*, ISTI, Bruxelles, 8-10 septembre 2005, p. 421-427. Disponible sur : <http://perso.univ-lyon2.fr/~thoiron/JS%20LTT%202005/pdf/Silva.pdf>

- [101] SINCLAIR, J., BALL, J. (1996). *EAGLES Preliminary Recommendations on Text Typology*. Birmingham University. EAG---TCWG---TTYP/P. June 1996. Disponible sur : <http://www.ilc.cnr.it/EAGLES/texttyp/texttyp.html>
- [102] SLODZIAN, M. (2000). « L'émergence d'une terminologie textuelle et le retour du sens. », In Bejoint H., Ph. Thoiron (dir.), *Le sens en terminologie*. Lyon : Presses Universitaires de Lyon, coll. « Travaux du C.R.T.T. », p. 61-85.
- [103] STAMATATOS, E., FAKOTAKIS, N., KOKKINAKIS, G. (2000). « Text genre detection using common word frequencies. » In *Proceedings of the 18th International Conference on Computational Linguistics, Association for Computational Linguistics*, vol. 2, Luxembourg, p. 808–814.
- [104] SUGAR BOESE E. (2005). « Stereotyping the Web: Genre Classification of Web documents. » *Master's Thesis*. Computer Science Department, Colorado State University. Disponible sur : <http://www.cs.colostate.edu/~boese/Research/masters.pdf>
- [105] TURENNE, N. (2000). « Apprentissage statistique pour l'extraction de concepts à partir de textes. Application au filtrage d'informations textuelles. » *Thèse de Doctorat en Informatique*. Université Louis-Pasteur Strasbourg/ENSAIS. Disponible sur : http://genome.jouy.inra.fr/~turenne/com_files/TheseTurenne.pdf
- [106] VAN DIJK, T.A. (1980). *Macrostructures*. Hillsdale, N.J.: Lawrence Erlbaum Associates, Publ.
- [107] VERONIS, J. (2003). « HyperLex : Cartographie lexicale pour la recherche d'informations. » *Rapport interne*. Equipe DELIC, Université de Provence. Disponible sur : <http://www.up.univ-mrs.fr/veronis/pdf/2003-hyperlex-rapport.pdf>
- [108] WERLICH, E. (1976). *A Text Grammar of English*. Heidelberg: Quelle & Meyer.
- [109] ZWEIGENBAUM, P., HABERT, B. (2006). « Faire se rencontrer les parallèles : regards croisés sur l'acquisition lexicale monolingue et multilingue. » *GLOTTOPOL* n. 8 « Traitements automatisés des corpus spécialisés : contextes et sens. » Juillet 2006. Disponible sur : http://www.univ-roen.fr/dyalang/glottopol/telecharger/numero_8/gpl8_03zweigenbaum_habert.pdf
- [110] ZIMINA, M. (2005). « Topographie bi-textuelle et approches quantitatives de l'extraction de ressources traductionnelles à partir de corpus parallèles », » In *Actes des 7es*

Journées scientifiques du Réseau de chercheurs 'Lexicologie, Terminologie, Traduction',
Bruxelles, 8-10 septembre 2005, p. 175-186. Disponible sur : <http://perso.univ-lyon2.fr/~thoiron/JS%20LTT%202005/pdf/Zimina.pdf>