

Ajuster corpus et objectifs – Éléments de bibliographie

Serge Fleury & Benoît Habert

<http://www.cavi.univ-paris3.fr/ilpga/ilpga/sfleury/> - fleurys [@] lim
<http://www.limsi.fr/Individu/habert> - habert [@] lim

29 octobre 2004

L'évolution rapide des données et des outils rend pratiquement obsolète un inventaire « papier » dès sa parution. On se contentera donc de renvoyer aux points d'entrée suivants :

- pour les ressources, les sites des deux fournisseurs principaux : LDC (*Linguistic Data Consortium* – <http://www ldc.upenn.edu/>) et ELRA (*European Language Resources Association* – <http://www.elra.info/>);
- pour les associations savantes :
 - ACL (Association for Computational Linguistics – <http://www.aclweb.org/>);
 - ATALA (Association pour le traitement automatique des langues – <http://www.atala.org/>);
 - AFCP Association Francophone de la Communication Parlée – <http://www.afcp-parole.org/>)
- pour les revues :
 - *Computational Linguistics* (voir le site de l'ACL);
 - *TAL (Traitement Automatique des langues)*; <http://www.atala.org/tal/tal.html> ;
 - le site d'articles sous formes électroniques lanl.arXiv.org, dans la rubrique *Computation and language* (<http://xxx.lanl.gov/list/cs.CL/recent>);
- pour un suivi des débats autour des ressources et de leur utilisation, les listes de discussion modérées suivantes :
 - Corpora (<http://helmer.aksis.uib.no/corpora/>);
 - LN-FR (<http://www.biomath.jussieu.fr/LN/LN-FR/>).

[Auroux, 1998]
[Barkema, 1993]
[Barkema, 1994]
[Beaudouin et al., 2002]
[Beauvisage, 2001]
[Biber et al., 1998]
[Biber et al., 1999]
[Bourigault and Fabre, 2000]
[Corbin, 1980]
[Fillmore, 1992]
[Firth, 1957]
[Folch et al., 2000]
[Gasiglia, 2004]
[Gross, 1988]
[Gross, 1996]
[Habert, 2000]
[Habert et al., 1998]
[Habert and Zweigenbaum, 2002]
[Harris et al., 1989]
[Harris, 1988]
[Harris, 1991]
[Illouz et al., 1999]
[Kilgarriff and Grefenstette, 2003]
[Kleiber, 1987]
[Leroy, 2004]
[Malrieu and Rastier, 2001]
[Manning and Schütze, 1999]
[Manning, 2003]
[Martin, 2001]
[Mela, 2004]
[Milner, 1989]
[Péchoin, 1992]
[Pincemin, 1999]
[Péry-Woodley, 1995]
[Reboul, 1994, p. 374]
[Rey, 1977]
[Simonin-Grumbach, 1975]
[Sinclair, 1996]
[Vaguer, 2004]
[Véronis, 2004]

Références

- [Auroux, 1998] Auroux, S. (1998). *La raison, le langage et les normes*. Sciences, modernités, philosophies. Presses Universitaires de France, Paris.
- [Barkema, 1993] Barkema, H. (1993). Idiomaticity in English NPs. In Aarts, J., de Haan, P., and Oostdijk, N., editors, *English language corpora : design, analysis and exploitation*, pages 257–278. Rodopi, Amsterdam.
- [Barkema, 1994] Barkema, H. (1994). Determining the syntactic flexibility of idioms. In Fries, U., Tottie, G., and Schneider, P., editors, *Creating and using English language corpora*, pages 39–52. Rodopi, Amsterdam.
- [Beaudouin et al., 2002] Beaudouin, V., Fleury, S., Habert, B., Pasquier, M., and Licoppe, C. (2002). Décrire la Toile pour mieux comprendre les parcours. Sites personnels et sites marchands. *Réseaux*, 20(116) :19–51. Parcours sur Internet. Valérie Beaudouin, Christian Licoppe (ed.).
- [Beauvisage, 2001] Beauvisage, T. (2001). Morpho-syntaxe et genres textuels. *TAL*, 42(2) :579–608. Linguistique de corpus – Béatrice Daille et Laurent Romary (resp.).
- [Biber et al., 1998] Biber, D., Conrad, S., and Reppen, R. (1998). *Corpus Linguistics : Investigating Language Structure and Use*. Cambridge Approaches to Linguistics. Cambridge University Press, Cambridge.
- [Biber et al., 1999] Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Longman, Harlow, Essex.
- [Bourigault and Fabre, 2000] Bourigault, D. and Fabre, C. (2000). Approche linguistique pour l’analyse syntaxique de corpus. *Cahiers de grammaire*, (25) :131–151.
- [Corbin, 1980] Corbin, P. (1980). De la production des données en linguistique introspective. In Dessaux-Berthoneau, A.-M., editor, *Théories linguistiques et traditions grammaticales*, pages 121–179. Presses Universitaires de Lille, Villeneuve-d’Asq.
- [Fillmore, 1992] Fillmore, C. J. (1992). « corpus linguistics » or « computer-aided armchair linguistics ». In Svartvik, J., editor, *Directions in Corpus Linguistics*, number 65 in Trends in Linguistics, pages 35–59. Mouton de Gruyter, Berlin. Proceedings of Nobel Symposium 82, Stockholm, 4–8 August 1991.
- [Firth, 1957] Firth, J. (1957). A synopsis of linguistic theory 1930–1955. *Studies in Linguistic Analysis*, pages 82–95. Réédité, *Selected Papers of J. R. Firth*, F. Palmer (edt), Longman.
- [Folch et al., 2000] Folch, H., Heiden, S., Habert, B., Fleury, S., Illouz, G., Lafon, P., Nioche, J., and Prévost, S. (2000). TyPTex : Inductive typological text classification analysis for NLP systems tuning/evaluation. In Maria Gavrilidou, George Carayannis, S. M. S. P. G. S., editor, *Second International Conference on Language Resources and Evaluation*, volume 1, pages 141–148, Athens, Greece. ELRA – European Language Resources Association.
- [Gasiglia, 2004] Gasiglia, N. (2004). Faire coopérer deux concordanciers-analyseurs pour optimiser les extractions en corpus. *Revue Française de Linguistique Appliquée*, IX(1) :45–62. « Linguistique et informatique : nouveaux défis » B. Habert (resp.).

- [Gross, 1988] Gross, G. (1988). Degré de figement des noms composés. *Langages*, (90) :57–70.
- [Gross, 1996] Gross, G. (1996). *Les expressions figées en français. Noms composés et autres locutions*. L'essentiel français. Ophrys, Paris.
- [Habert, 2000] Habert, B. (2000). Review of *an introduction to corpus linguistics*, by Graeme Kennedy, Longman, 1998. *Journal of Quantitative Linguistics*, 7(3) :245–250.
- [Habert et al., 1998] Habert, B., Fabre, C., and Issac, F. (1998). *De l'écrit au numérique : constituer, normaliser, exploiter les corpus électroniques*. Informatiques. InterÉditions/Masson, Paris.
- [Habert and Zweigenbaum, 2002] Habert, B. and Zweigenbaum, P. (2002). Régler les règles. *TAL*, 43(3) :83–105. Problèmes épistémologiques. M. Cori, S. David, J. Léon (resp.).
- [Harris, 1988] Harris, Z. (1988). *Language and information*. Columbia University Press, New York.
- [Harris et al., 1989] Harris, Z., Gottfried, M., Ryckman, T., Mattick Jr, P., Daladier, A., Harris, T., and Harris, S. (1989). *The Form of Information in Science, Analysis of Immunology Sublanguage*, volume 104 of *Boston Studies in the Philosophy of Science*. Kluwer Academic Publisher, Dordrecht, The Netherlands.
- [Harris, 1991] Harris, Z. S. (1991). *A theory of language and information. A mathematical approach*. Oxford University Press, Oxford.
- [Illouz et al., 1999] Illouz, G., Habert, B., Fleury, S., Folch, H., Heiden, S., and Lafon, P. (1999). Maîtriser les déluges de données hétérogènes. In Condamines, A., Fabre, C., and Péry-Woodley, M.-P., editors, *Corpus et traitement automatique des langues : pour une réflexion méthodologique*, pages 37–46, Cargèse.
- [Kilgarriff and Grefenstette, 2003] Kilgarriff, A. and Grefenstette, G. (2003). Introduction to the special issue on the web as corpus. *Computational Linguistics*, 29(3) :335–347.
- [Kleiber, 1987] Kleiber, G. (1987). Quelques réflexions sur le vague dans les langues naturelles. In *Etudes de linguistique générale et de linguistique latine offertes à Guy Serbat*, pages 157–172. Bibliothèque de l'information, Paris.
- [Leroy, 2004] Leroy, S. (2004). Extraire sur patrons : allers et retours entre analyse linguistique et repérage automatique. *Revue Française de Linguistique Appliquée*, IX(1) :25–43. « Linguistique et informatique : nouveaux défis » B. Habert (resp.).
- [Malrieu and Rastier, 2001] Malrieu, D. and Rastier, F. (2001). Genres et variations morphosyntaxiques. *TAL*, 42(2) :547–577. Linguistique de corpus – Béatrice Daille et Laurent Romary (resp.).
- [Manning, 2003] Manning, C. D. (2003). Probabilistic syntax. In Bod, R., Hay, J., and Jannedy, S., editors, *Probabilistic Linguistics*, pages 289–341. The MIT Press, Cambridge, Massachusetts.
- [Manning and Schütze, 1999] Manning, C. D. and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, Massachusetts.

- [Martin, 2001] Martin, R. (2001). *Sémantique et automate*. Écritures électroniques. Presses Universitaires de France, Paris.
- [Mela, 2004] Mela, A. (2004). Linguistes et « talistes » peuvent coopérer : repérage et analyse des gloses. *Revue Française de Linguistique Appliquée*, IX(1) :63–82. « Linguistique et informatique : nouveaux défis » B. Habert (resp.).
- [Milner, 1989] Milner, J.-C. (1989). *Introduction à une science du langage*. Des Travaux. Seuil, Paris, 1ère édition.
- [Péchoin, 1992] Péchoin, D., editor (1992). *Thésaurus Larousse – des idées aux mots, des mots aux idées*. Larousse, Paris, 2ème édition.
- [Pincemin, 1999] Pincemin, B. (1999). Sémantique interprétative et analyse automatique de textes : que deviennent les sèmes ? *Sémiotiques*, (17) :71–120. *Dépasser les sens iniques dans l'accès automatisé aux textes*, B. Habert (resp.).
- [Péry-Woodley, 1995] Péry-Woodley, M.-P. (1995). Quels corpus pour quels traitements automatiques ? *TAL*, 36(1–2) :213–232. *Traitements probabilistes et corpus*, Benoît Habert (resp.).
- [Reboul, 1994] Reboul, A. (1994). Concepts flous et usages approximatifs. In Moeschler, J. and Reboul, A., editors, *Dictionnaire encyclopédique de pragmatique*, chapter 14, pages 373–397. Seuil, Paris.
- [Rey, 1977] Rey, A. (1977). *Le lexique : images et modèles, du dictionnaire à la lexicologie*. Linguistique. Armand Colin, Paris.
- [Simonin-Grumbach, 1975] Simonin-Grumbach, J. (1975). Pour une typologie des discours. In *Langue, discours, société (pour Emile Benveniste)*, pages 85–121. Seuil, Paris.
- [Sinclair, 1996] Sinclair, J. (1996). Preliminary recommendations on corpus typology. Technical report, EAGLES (Expert Advisory Group on Language Engineering Standards).
- [Vaguer, 2004] Vaguer, C. (2004). Constitution d'une base de données : les emplois de *dans* marquant la coïncidence. *Revue Française de Linguistique Appliquée*, IX(1) :83–97. « Linguistique et informatique : nouveaux défis » B. Habert (resp.).
- [Véronis, 2004] Véronis, J. (2004). Quels dictionnaires pour l'étiquetage sémantique ? *Le français moderne*, 72(1) :27–38. *Traitement automatique et ressources numérisées pour le français*. C. Fuchs et B. Habert (resp.).