

| | |
|--|---|
| 9. Kongress des Frankoromanistenverbands 24. bis 27. September 2014 Universität Münster Thema: „Schnittstellen/Interfaces“ | IXe congrès de l'Association des francoromanistes 24-27 septembre 2014 Université de Münster Thématique « Schnittstellen/Interfaces » |
|--|---|

Atelier 2: Les “interfaces numériques” dans la recherche d'aujourd'hui

Organisation:

Prof. Dr. Elisabeth Burr (Leipzig; elisabeth.burr@uni-leipzig.de)

Dr. Christof Schöch (Würzburg; christof.schoech@uni-wuerzburg.de)

Exposé de l'atelier

Les chercheurs et chercheuses dans le domaine de la langue, de la littérature et de la culture françaises et francophones peuvent de plus en plus utiliser, dans leurs travaux, des représentations numériques de leurs objets de recherche. En fonction des thématiques, des époques et des approches méthodiques, des matériaux de forme et de type différents peuvent être pertinents : les facsimilés numériques de documents, les grandes collections de textes littéraires sous forme de texte simple, les éditions critiques en ligne des oeuvres d'un auteur donnée, ou bien les corpus linguistiques plus ou moins fortement annotés. Ces nouveaux objets de recherche impliquent également que de nouvelles méthodes de recherches soient développées et mises en oeuvre à l'aide d'outils divers.

Dans ce contexte, nous proposons de reprendre à Milhad Doueïhi le concept des “interfaces numériques” et de le concrétiser dans le contexte de la recherche en sciences humaines.

Par exemple, les représentations numériques des objets de recherche peuvent eux-mêmes être considérées comme des “interfaces” positionnées entre les chercheurs et chercheuses, d'une part, et le phénomène culturel, littéraire ou linguistique qui motive leur recherche, d'autre part. Par ailleurs, les outils de découverte, d'analyse, de lecture et de manipulation des données numériques présentent également des “interfaces” aux chercheurs et chercheuses qui les utilisent. Enfin, les méthodes numériques fonctionnent comme des “interfaces” ou des lieux de rencontre entre des chercheurs et chercheuses en linguistique, en études de lettres, en science de l'information et en informatique qui se réunissent autour d'un problème ou d'une méthode de recherche.

L'atelier se propose de réunir autour de la question des “interfaces numériques” des chercheurs et chercheuses venant de disciplines différentes, comme la linguistique, les études de lettres et l'informatique, et qui s'intéressent aux interfaces numériques de recherche dans des perspectives différentes. Parmi les questions qui pourront être adressées par les participants et participantes de l'atelier, on peut nommer les suivantes, à titre d'exemple et en allant du général au spécifique :

- Qu'est-ce que les méthodes, données et outils numériques apportent-ils à des domaines de recherche établis, tels que l'histoire littéraire, la stylistique, l'intertextualité, la recherche sur la constitution de “gender”, sur l'usage de la langue, etc.?
- Quelles configurations interdisciplinaires, quels fonctionnements nouveaux les “interfaces numériques” produisent-elles dans la recherche en linguistique, littérature ou culture?

- Comment les savoirs et les compétences que littéraires, linguistes et informaticiens ou informaticiennes apportent à un projet de recherche sont-ils mis à contribution et coordonnées concrètement?
- Comment la découverte des ressources et des textes, la communication et la collaboration, enfin la dissémination des résultats de la recherche sont-elles modifiées par le paradigme numérique?
- Comment des effets de décanonisation, d'élargissement des corpus, de collaboration à grande échelle modifient-ils la recherche dans les études de genre?
- Quels sont les avantages et les difficultés concernant l'interface proposé aux utilisateurs et utilisatrices d'une édition de texte numérique?
- Comment la question des variantes textuelles peut-elle être adressée avec des outils numériques et des visualisations?

Nous voudrions proposer d'organiser, en complément aux travaux présentés dans les sessions de l'atelier, une présentation des projets de recherche menés par les participants et participantes sous forme d'affiches dans un espace ouvert à tous les participants et toutes les participantes au colloque.

Références (sélection)

David M. Berry (ed.) (2012): *Understanding Digital Humanities*. Houndmills, Basingstoke: Palgrave Macmillan.

Christine Borgman (2011): *Scholarship in the Digital Age. Information, Infrastructure and the Internet* (2007). Cambridge: MIT Press.

Milad Doueïhi (2012): *Pour un humanisme numérique* (2011). Paris: publie.net.

Stephen Ramsay (2011): *Reading Machines. Toward an Algorithmic Criticism*, Urbana: Univ. of Illinois Press.

Sinclair, John (1991): *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Meunier, Fanny / De Cock, Sylvie / Gilquin, Gaëtanelle / Paquot, Magali (eds.) (2011): *A Taste for Corpora. In honour of Sylviane Granger* (= Studies in Corpus Linguistics 45). Amsterdam: John Benjamins.

Plan des communications

Jeudi, 25.09.2014

| | |
|-------------|---|
| 9:00-9:45 | Introduction à l'atelier |
| 9:45-10:30 | Sascha Diwersy: Vers une linguistique variationnelle outillée - La base <i>Varitext</i> et le <i>Corpus des variétés nationales du français</i> |
| 10:30-11:00 | Kaffeepause |
| 11:00-12:00 | Plenarvortrag Hörsaal |
| 12:30-14:30 | Mittagspause |
| 14:30-15:15 | Julia Burkhardt, Aramis C. Durán, Elena Potapenko, Rebecca Sierig: FrltZ Le corpus de la langue des journaux français et italiens de Leipzig - Développement et possibilité d'application d'un corpus de la presse écrite |
| 15:15-16:00 | Olga Kellert: Constructions avec c'est : une recherche dans les corpus |
| 16:00-16:30 | Pause |
| 16:30-17:15 | Dirk Siepmann, Christoph Bürgel: Das Corpus de Référence du français contemporain |
| 17:15-18:00 | Fleur Pfeifer: La communication scientifique ou: l'essai d'une méta-analyse |
| 18:00-20:00 | Mitgliederversammlung Hörsaal |
| ab 20:00 | Lesung Hörsaal |

Vendredi, 26.09.2014

| | |
|-------------|---|
| 9:00-9:45 | Marie-Hélène Lay & Sascha Diwersy: Elaboration assistée de modèles d'annotation : protocole et interface numérique |
| 9:45-10:30 | David Wrisley: Lieu, Temps, Réseau : la modélisation numérique des « géographies littéraires » du moyen âge français |
| 10:30-11:00 | Kaffeepause |
| 11:00-12:00 | Plenarvortrag Hörsaal |
| 12:30-14:30 | Mittagspause |
| 14:30-15:15 | Marianne Le Galliard: Quand la maîtrise de l'outil informatique va de pair avec un projet de recherche sur la photographie de Jacques Henri Lartigue |

| | |
|-------------|---|
| 15:15-16:00 | Christof Schöch: L'histoire du genre policier à travers ses "topics" |
| 16:00-16:30 | Pause |
| 16:30-17:15 | Suzan van Dijk: Collaborations rendues possibles grâce à Internet |
| 17:15-18:00 | Lars Schneider, Annette Gerstenberg & Christof Schöch: La plate-forme romanistik.de comme interface numérique |

Samedi, 27.09.2014

| | |
|-------------|--|
| 9:00-9:45 | Mokhtar Ben Henda & Henri Hudrisier: TEI et humanités digitales en milieu linguistique franco-maghrébin |
| 9:45-10:30 | Thomas Bremer: Philologische Überlegungen zu einem elektronischen Editionsprojekt: Raynals <i>Histoire des deux Indes</i> im semiautomatischen Textvergleich |
| 10:30-11:00 | Kaffeepause |
| 11:00-11:45 | Susanne Schütz / Marcus Pöckelmann: IT-Werkzeuge zur Unterstützung elektronischer Edition am Beispiel eines französischen Textes aus dem 18. Jahrhundert |
| 11:45-12:30 | Discussion d'ensemble et bilan |
| 12:30-14:30 | Mittagspause |

Exposés

Thomas Bremer (Halle)

Philologische Überlegungen zu einem elektronischen Editionsprojekt: Raynals *Histoire des deux Indes* im semiautomatischen Textvergleich.

Die *Histoire des deux Indes* von Thomas Guillaume Raynal (1713-1796), ein zentraler Text der französischen Aufklärung, liegt in vier gedruckten Fassungen (1770, 1776, 1780, posthum 1824) vor, die über einen Zeitraum von mehr als 20 Jahren vom Autor überarbeitet und erweitert wurden. Der Text ist die erste politisch-historische Abrechnung mit der europäischen Kolonialpolitik und zentriert sich auf die Karibik und Lateinamerika („Westindien“) einerseits und auf die asiatischen („Ostindien“) Kolonialräume andererseits, wurde aber jeweils sofort nach Erscheinen in Frankreich verboten, sodass die Ausgaben jeweils außerhalb Frankreichs (Genf, Den Haag) erschienen.

Im hier vorgestellten Projekt wird die Textgenese am Beispiel der Lateinamerikadarstellung in der *Histoire* v.a. unter buchhistorischem Gesichtspunkt nachvollzogen. Es lässt sich zeigen, dass die Überarbeitungen (die von der Korrektur von Schreib- und Druckfehlern bis hin zu seitenlangen Ergänzungen des Ursprungstextes alle denkbaren Formen von Varianten umfassen) zu einer zunehmenden Radikalisierung in Raynals Denken und den Zuwachs an Wissen über Lateinamerika im ausgehenden 18. Jahrhundert führen. Der Vortrag zeigt in diesem Sinne den Nutzen einer genetischen Textedition, der die jeweiligen Erweiterungen auf der Grundlage des Ursprungstexts deutlich macht.

Weitere Informationen:

<http://informatik.uni-halle.de/ti/forschung/ehumanities/sada>

Julia Burkhardt / Aramis C. Durán / Elena Potapenko / Rebecca Sierig (Leipzig)

FrltZ: Le corpus de la langue des journaux français et italiens de Leipzig - Développement et possibilité d'application d'un corpus de la presse écrite

Les corpus digitaux ont de plus en plus d'importance quant aux recherches linguistiques actuelles. Le travail sur corpus rend possible des résultats quantitatifs et qualitatifs sur la parole inaccessibles par l'analyse manuelle. Pourtant, il y a toujours peu de corpus accessibles qui représentent les variétés de l'italien ou du français. Le présent projet, associé à la chaire de linguistique italienne et française à Leipzig, s'adonne au développement d'un corpus de la langue des journaux français et italiens.

La langue des journaux sera définie comme variété de langue écrite qui est vue selon Burr (1993 : 141) comme « faisceau de styles » qui est dans sa totalité adapté au médium : il s'adapte respectivement aux types de textes divers, à la gamme de thèmes et domaines dedans et enfin aux émetteurs de même qu'aux lectrices. La langue des journaux est un phénomène complexe : elle obéit à des normes collectives tout en exerçant son influence sur celles-ci et elle unit des styles variés. Elle est ainsi une interface au sens propre de notre congrès.

Le corpus FrltZ (Korpus Französischer und Italienischer Zeitungssprachen) se distinguera d'autres corpus de presse par le fait qu'il prend au sérieux cette conception globale des langues des journaux dans laquelle s'imbriquent tous les éléments constitutifs du journal. La conception du corpus tient en compte cette imbrication : des exemplaires complets de quotidiens français et italiens sont insérés dans le corpus et ils sont pourvus d'annotations abondantes. Le markup du corpus s'oriente aux *Guidelines* de la TEI en liant le corpus à la

recherche internationale et interdisciplinaire. Un but du projet est au surplus l'intégration du *Corpus des Langues de journaux Romans* (Burr 1997-2004) crée auparavant et déjà disponible en ligne. Finalement, le corpus sera rendu accessible dans son intégralité à la communauté scientifique. Cette contribution vise à présenter la conception et la préparation du corpus comme 'interface numérique'. Enfin, les possibilités d'utilisation du corpus seront esquissées de manière exemplaire.

Références

Burr, Elisabeth (1993): *Verb und Varietät. Ein Beitrag zur Bestimmung der sprachlichen Variation am Beispiel der italienischen Zeitungssprache*. Hildesheim: Olms.

Burr, Elisabeth (1997-2004): *Korpus Romanischer Zeitungssprachen*. Duisburg / Bremen / Leipzig. <http://www.uni-leipzig.de/~burr/CorpusLing/> [23.01.2014].

Sascha Diwersy (Köln)

Vers une linguistique variationnelle outillée - La base *Varitext* et le *Corpus des variétés nationales du français*

La prise en compte des données quantitatives dans l'étude du diasystème fait l'objet d'un intérêt croissant. Le dépouillement automatique de grands atlas linguistiques, comme le propose la dialectométrie (Goebel & Rézeau 1995) a ainsi apporté des éclairages nouveaux pour l'identification des variétés diatopiques. Ces méthodes prolongent le statut intrinsèquement quantitatif de la notion de faisceaux d'isoglosses. Du côté des grands corpus textuels, on peut enfin prendre en compte l'incidence des genres textuels et des traditions discursives (Schlieben-Lange 1983) sur la diffusion des innovations et l'interaction entre les niveaux de variation. Ainsi, Glessgen (2007 : 424) peut-il avancer que « les méthodes de quantification sont sur le point de transformer la recherche et l'histoire linguistiques comme elles ont transformé, il y a quelques décennies, la recherche sociologique ou psychologique ».

Notre communication se proposera de présenter le *Corpus des variétés nationales du français* (COVANA-FR), une banque de données textuelles portant sur la variation du français écrit dans l'espace francophone. Les archives numériques en question comprennent actuellement plusieurs quotidiens en provenance de dix pays (Algérie, Cameroun, Côte d'Ivoire, France, Mali, Maroc, RDC, Sénégal, Suisse, Tunisie), pour un total de 350.000.000 de mots. Elles sont interrogeables par l'intermédiaire de la plate-forme d'analyse lexico-statistique liée à la base *Varitext* (Diversy, Meiri & Blumenthal 2013), qui est accessible gratuitement en ligne.

L'ensemble des documents disponibles est doté, d'une part, de métadonnées permettant de les regrouper en termes d'origine géographique, de dates et de thématiques. D'autre part, les textes concernés ont fait l'objet d'une lemmatisation, d'une catégorisation morpho-syntaxique et d'un étiquetage au niveau des relations de dépendance.

Le corpus ainsi constitué est adossé à des méthodologies d'analyses à dominante inductive (angl. corpus-driven) intéressant également la linguistique variationnelle. Ces méthodes, dans la lignée du concept des « profils combinatoires » (cf. Blumenthal 2006a, 2006b ; Diwersy 2012a) ainsi que de la Théorie de l'amorçage lexical (Lexical Priming Theory, cf. Hoey 2005), ont pour objectif de mettre au jour des spécificités d'usage liées à des patrons co-occurrence divergents tant au niveau lexical que syntaxique.

A cet égard, le logiciel d'exploration associé au COVANA-FR intègre plusieurs applications, par lesquelles sont mis en place les procédés lexico-statistiques suivants : calculs de spécificités fréquentielles (Lafon 1980) et de « mots clés » (Rayson 2003), détermination d'affinités combinatoires (collocationnelles et colligationnelles) au moyen de certaines

mesures d'associations telles que le log-likelihood ratio (cf. Dunning 1993), et regroupement des données co-occurrenceelles sous forme de listes d'items attirés de façon préférentielle, c'est-à-dire sous forme de lexicogrammes (cf. Heiden 2004), exploration des lexicogrammes en termes de symétries cooccurrenceelles (cf. Luong et al. 2010), modélisation des (dis)similarités entre lexicogrammes au moyen de méthodes multivariées telles que l'échellonnement multidimensionnel (cf. Blumenthal 2009), le partitionnement de données (Classification ascendante hiérarchique, cf. Diwersy 2012b) ou l'Analyse factorielle des correspondances (cf. Diwersy & Loiseau à paraître).

Dans notre exposé, nous nous concentrerons sur une partie des applications mentionnées en illustrant leur fonctionnement par des requêtes exemplaires destinées à faire ressortir, dans une perspective variationnelle, les spécificités combinatoires d'un certain nombre d'items lexicaux.

Références

- Blumenthal, P. (2006a): *Wortprofil im Französischen*. Tübingen: Niemeyer.
- Blumenthal, P. (2006b): « De la logique des mots à l'analyse de la synonymie ». *Langue française* 150, 14-31.
- Blumenthal, P. (2009): « Les noms d'émotion : trois systèmes d'ordre ». In: Novakova, I. & Tutin, A. (Eds): *Le Lexique des émotions*. Grenoble: ELLUG (Langues, Gestes, Paroles), 41-64.
- Blumenthal, P. & Diwersy, S. & Mielebacher, J. (2005): « Kombinatorische Wortprofile und Profilkontraste: Berechnungsverfahren und Anwendungen ». *Zeitschrift für romanische Philologie* 121, 49-83.
- Diwersy, S. (2012a): *Kookkurrenz, Kontrast, Profil: Korpusinduzierte Studien zur lexikalisch-syntaktischen Kombinatorik französischer Substantive (mit ergänzenden Betrachtungen zum Deutschen)*. Berlin e.a.: de Gruyter.
- Diwersy, S. (2012b): « La francophonie multivariée ou Comment mesurer les français en Afrique ? ». *Le français en Afrique* 27, 75-91.
- Diwersy, S. & Mejri, S. & Blumenthal, P. (2013): *Base Varitext*. Köln; Villetaneuse : Romanisches Seminar; LDI (UMR 7187), <http://syrah.uni-koeln.de/varitext/> , <http://extranet-ldi.univ-paris13.fr/varitext/> .
- Diwersy, S. & Loiseau, S. (à paraître): « La différenciation du français dans l'espace francophone : l'apport des statistiques lexicales ».
- Dunning, T. (1993): « Accurate Methods for the Statistics of Surprise and Coincidence ». *Computational Linguistics* 19, 1, 61-74.
- Glessgen, M.-D. (2007): *Linguistique romane, domaine et méthode – Domaines et méthodes en linguistique française et romane*. Paris: Armand Colin.
- Goebel, H. & Rézeau, P. (1995) : « L'interprétation de certes dialectométriques », *RLiR* 69, 297-309.
- Heiden, S. (2004): « Interface hypertextuelle à un espace de cooccurrences: implémentation dans Weblex ». In : Purnelle, G. & Fairon, C. & Dister, A. (Eds.), *Le poids des mots: Actes des 7es Journées internationales d'Analyse statistique des Données Textuelles (JADT 04, Louvain-la-Neuve, 10-12 mars 2004)*. Louvain-la-Neuve : Presses Universitaires de Louvain, 577-588.
- Hoey, M. (2005): *Lexical Priming: a new theory of words and language*. London: Routledge.
- Lafon, P. (1980): « Sur la variabilité de la fréquence des formes dans un corpus ». *Mots* 1, 127-165.
- Luong, X. & Brunet, E. & Longrée, D. & Mayaffre, D. & Mellet, S. & Poudat, C. (2010): « La cooccurrence, une relation asymétrique ? ». In: Bolasco, S. & Chiari, I. & Giuliano, L. (Eds.): *Jadt 2010 - Statistical Analysis of Textual Data. Proceedings of 10th International Conference Journées d'Analyse statistique des Données Textuelles, 9-11 June 2010 - Sapienza University of Rome*. Milan : Edizioni Universitarie di Lettere Economia Diritto, 321-331.

Rayson, P. (2003): *Matrix: A statistical method and software tool for linguistic analysis through corpus comparison*. Thèse de doctorat, Lancaster University.
Schlieben-Lange, B. (1983) : *Traditionen des Sprechens*. Stuttgart: Kohlhammer.

Henri Hudrisier (Paris) / Mokhtar Ben Henda (Bordeaux) **TEI et humanités digitales en milieu linguistique franco-maghrébin**

Depuis des siècles, la région du Maghreb connaît une hybridation linguistique importante qui impacte lentement sur son patrimoine culturel. Des quantités non négligeables de ressources dans plusieurs cultures linguistiques (arabe classique, dialectes régionaux, dialectes berbères, français, anglais) se sont accumulés à travers les siècles. Notre hypothèse est que le seul fait de faire connaître cette diversité des ressources textuelles aurait pour effet d'augmenter de façon exponentielle le nombre de chercheurs susceptibles de s'intéresser à l'étude de ces littératures plurielles. Le format de codage standard que constitue la TEI, présente une opportunité unique, celle de créer ces ressources en garantissant leur intégration dans le patrimoine culturel général et leur réutilisabilité avec une flexibilité technique maximale.

Dans notre projet « HumanitéDigitMaghreb », nous partons de cette hypothèse pour étudier la faisabilité, et amorcer l'appropriation de méthodes, permettant de répondre à plusieurs aspects de ces objectifs de recherche. Le projet nous a permis surtout des visions comparées qui permettent de comprendre comment transformer la TEI pour travailler sur les corpus en langues et écritures hybrides. Dans le projet, nous nous sommes focalisés particulièrement sur les méthodes de mise en œuvre de la TEI pour traiter des cas spécifiques de la structuration de corpus multilingues. Nous donnerons ici quelques résultats de nos travaux, mais nous examinerons surtout les enjeux concrets et prospectifs de très grands corpus linguistiquement structurés de façon normative qui permettront de constituer pour toute communauté linguistique le référentiel de contexte indispensable pour interagir avec les outils de traductique et d'e-semantique du futur.

En nous inscrivant dans l'école de pensée des Humanités digitales et de la TEI, nous voulons explicitement donner non seulement une réalité tangible et numérique à nos travaux mais les rendre facilement échangeables, cumulables, améliorables, modifiables partout dans le monde. Nous voulons aussi sur un plan plus spécifique que nos travaux soient déjà facilement échangeables et cumulables entre nous et avec nos trois langues partenaires maghrébines (arabe, français, berbère) auxquelles il convient de rajouter l'anglais.

Bibliographie

- Abbès R. (2000). "Encodage des corpus de textes arabes en conformité à la TEI, outils et démarche technique". Rapport final de projet DIINAR-MBC.
- Burnard, L. (2012). "Encoder l'oral en TEI : démarches, avantages, défis.... Présenté à Conférence à la Bibliothèque Nationale de France, Paris: Abigaël Pesses.
- Guesdon, Marie-Genviève (2008). "Bibliothèque nationale de France: Manuscripts catalogue 'Archives et manuscrits'". Paper presented at the Fourth Islamic Manuscript Conference, Cambridge
- Ide, N. (1996). "Representation schemes for language data: the Text Encoding Initiative and its potential impact for encoding African languages". In CARI'96.
- Ide, N. M., & Véronis, J. (1995). *Text Encoding Initiative: Background and Contexts*. Springer.
- Jungen, C. (2012). "Quand le texte se fait matière". *Terrain*, n° 59 (2), 104-119.
- Ourabah Mohammed, S., & Hassoun, M. (2012). "A TEI P5 Manuscript Description Adaptation for Cataloguing Digitized Arabic Manuscripts". *Journal of the Text Encoding Initiative*, 2.

Olga Kellert (Göttingen)

Constructions avec c'est : une recherche dans les corpus

L'objectif de cette communication consiste à montrer ce que les corpus choisis apportent à la recherche de l'usage de quelques constructions avec c'est (voir ex. 1-2) (Lambrecht 1996, Clech-Darbon et al. 1999):

(1) C'est vous qui le justifiez.

(2) C'est à vous de le justifier.

Malgré les différences syntaxiques entre ex. 1 et 2 (comme p.ex. l'infinitif dans l'ex. 2 et verbe fini dans l'ex. 1), nous verrons sur la base des données détectées dans les corpus CORAL, COWFR2011 et Frantext que les deux types de constructions peuvent se ressembler en ce qui concerne leur fonction pragmatique. On les utilise souvent pour mettre en contraste l'expression qui suit la forme *c'est* (aussi Dassi 2003 : 247). Cela se traduit par la négation d'un élément associé à un ensemble d'alternatives (ici : toi, vous, etc.):

(3) Donc c'est à moi d'appeler euh les médias. (CORAL ffamdl 19).

(4) C'était à moi et rien qu'à moi de... (Dassi 2003 :247)

L'expression d'un contraste est souvent liée à la notion de focus (Dassi 2003 : 247, Doetjes et al. 2004, Hamlaoui 2007). Nous verrons que la fonction de focus reflète non seulement la fréquence des éléments de contraste dans l'énoncé qui suit la construction avec *c'est* (p.ex. pas), mais elle reflète encore l'usage de la forme intonatoire. La phrase qui suit *c'est* démontre une saillance prosodique à travers une courbe intonative (description de l'intonation dans l'ex. 1 dans Doetjes et al. 2004, Hamlaoui 2007).

Les corpus mettent ainsi en évidence l'usage des constructions en question dans des contextes authentiques (sur l'utilité des corpus pour la description de la langue, voir Sinclair 1991 entre autres).

Références

- Clech-Darbon, Anne et al. (1999): «Are There Cleft Sentences in French? », in: G. Rebuschi, L. Tuller (éds.). *The Gram-mar of Focus*. Amsterdam, 83-118.
- C-ORAL-ROM (CORAL) = Cresti, E. / Moneglia, M. (éds.) (2005): *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*, Amsterdam/ Philadelphia.
- Dassi, M. (2003). *De la morphosyntaxe à la sémantique des présentatifs en français contemporain. Une aperception fondée sur la prose romanesque de Mongo Beti*. München : Lincom.
- FRANTEXT = Base textuelle FRANTEXT, ATILF - CNRS & Université de Lorraine. Site internet: <http://www.frantext.fr>.
- FRCOW2011 <http://hpsg.fu-berlin.de/cow/colibri/>
- Lambrecht, Knud (1996): *Information structure and sentence form: topic focus and the mental representation of discourse referents*. Cambridge: Cambridge University Press.
- Sinclair, John (1991): *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Marie-Hélène Lay (Poitiers) / Sascha Diwersy (Köln)

Elaboration assistée de modèles d'annotation : protocole et interface numérique

L'annotation, définie comme de l'ajout d'information à de l'information disponible, consiste à associer des méta-données à des données (cf. Leech 1997, Habert 2005). C'est un processus indissociable de l'activité en sciences humaines : les « notes », étiquettes associées, en accompagnent toutes les étapes, depuis les premières prises de contacts avec les données, la formulation d'hypothèses jusqu'à l'élaboration de concepts, de

catégories jugées pertinentes pour représenter les données à un niveau « méta ». Celles-ci sont alors des indications qui rendent l'objet d'étude « préhensible » (cf. Houdé, 1992) et permettent de construire des filtres d'extraction mis en œuvre dans l'analyse des données. Dans le domaine du Traitement Automatique du Langage, les annotations sont souvent d'ordre morpho-syntaxique. Les interfaces numériques, quant à elles, peuvent accompagner toutes les étapes d'un tel processus (cf. Fort 2012), de l'élaboration des catégories à leur encodage sur des volumes plus ou moins importants de données.

Cette contribution en propose un exemple : ce dont nous faisons état ici est la mise en place d'un protocole réutilisable, permettant la création et la mise à disposition de corpus annotés, dont les usages ultérieurs sont multiples. Ici, le corpus généré est destiné à l'élaboration d'un modèle d'étiquetage pour le français pré-classique : le corpus annoté est utilisé dans les phases d'entraînement et de validation du tagger.

Nous articulerons la présentation de ce protocole à sa mise en œuvre par le biais d'un outil présentant une interface numérique : AnaLog. Il s'agit d'un outil permettant de poser les étiquettes de façon assistée, d'enrichir les ressources linguistiques (dictionnaires et corpus annotés) et de les adapter progressivement en fonction des décisions prises lors du déroulement de la campagne d'annotation (cf. Lay, 2010).

L'expérimentation elle-même porte sur la création d'une ressource (lexique de formes et corpus d'apprentissage) susceptible d'alimenter un modèle d'annotation automatique adapté au français du 17^e siècle. La motivation de cette tâche, menée dans le cadre du projet ANR-DFG Presto (Il s'agit d'un projet franco-allemand traitant de l'évolution diachronique du système prépositionnel en français et animé conjointement par le laboratoire Icar à l'ENS de Lyon - responsable : D. Vigier - ainsi que par l'Institut des Langues romanes à l'Université de Cologne - responsable : P. Blumenthal.) tient au fait qu'il n'existe pas de modèle librement disponible pour le français (pré)classique, contrairement à ce qui se passe pour le français ancien et le français contemporain (Il existe des modèles du catégoriseur TreeTagger pour l'ancien français et pour le français contemporain (cf. Stein 2003, Stein, Schmid 1995) ; Parmi les autres modèles disponibles pour le français contemporain on peut citer ceux qui sont utilisés par MElt (Denis, Sagot 2009), morfette (Chrupala et al. 2008), LGTagger (Constant, Sigogne 2011) et Talismane (Urieli, Tanguy 2013).). Ce modèle a la particularité de pouvoir être paramétré afin de prendre en compte de façon efficace la variation des formes observables du début du 17^e à la fin du 18^e : on pourra ainsi rendre compte du passage d'un état du français à l'autre.

Références

- Crabbé B., Candito M. (2008). Expériences d'analyse syntaxique statistique du français. In *Actes de la 15^{ème} Conférence sur le Traitement Automatique des Langues Naturelles (TALN'08)*, Avignon, France, 45–54.
- Denis P., Sagot S. (2009). Coupling an annotated corpus and a morphosyntactic lexicon for state-of-the-art POS tagging with less human effort. In *Proceedings of The Pacific Asia Conference on Language, Information and Computation (PACLIC 23)*, Hong Kong, Chine.
- Fort, K. (2012). Les ressources annotées, un enjeu pour l'analyse de contenu : vers une méthodologie de l'annotation manuelle de corpus, Ph.D. Université Paris 13.
- Habert, B. (2005). *Instruments et ressources électroniques pour le français*. Paris /Gap : Ophrys.
- Houdé O. (1992). *Catégorisation et développement cognitif*, Paris : PUF.
- Lay M.-H., Demonet M.-L. (2000). Adaptation d'un lemmatiseur au corpus rabelaisien, naissance d'Humanistica. In *Actes de JADT*, EPFL, 267-278.
- Lay M.-H. (2010). Pour une exploration humaniste des textes : AnaLog . In *Actes de JADT 2010*, Roma, 1045-1056.
- Leech G. (1997). *Corpus annotation : Linguistic information from computer text corpora, chapitre Introducing corpus annotation*, Londres : Longman, 1–18.

Padró L., Stanilovsky E. (2012). FreeLing 3.0: Towards Wider Multilinguality, in *Proceedings of the Language Resources and Evaluation Conference (LREC 2012) ELRA*. Istanbul, Turkey. May, 2012.

Romary L., Salmon-Alt S., Francopoulo G. (2004). Standards going concrete : from LMF to Morphalou. *Workshop on Electronic Dictionaries, Coling 2004*, Geneva, Switzerland.

Sagot B. (2010). The Lefff, a freely available and large-coverage morphological and syntactic lexicon for French. In *Proceedings of the 7th international conference on Language Resources and Evaluation (LREC 2010)*, Istanbul, Turquie.

Martine le Galliard (Paris)

Quand la maîtrise de l'outil informatique va de pair avec un projet de recherche sur la photographie de Jacques Henri Lartigue

La base de données sur l'œuvre photographique de Jacques Henri Lartigue développée depuis 2003 a permis récemment de changer de point de vue sur les albums de famille construits de 1890 à 1986. L'outil numérique créant des liens entre les images (pages d'albums, tirages photo, dessins) et archives (correspondance, notes, journal intime) a été entièrement conceptualisé et construit par un chercheur-développeur, maîtrisant les code xml et html et programmant la base en fonction de ses propres besoins scientifiques. La saisie, les numérisations, la création des liens, la mise en forme des affichages, les indexations ont été entièrement effectuées par celui-ci. Les découvertes ont fait l'objet d'une thèse universitaire intitulée « Jacques Henri Lartigue sous le regard de Richard Avedon. La question de l'album photographique à partir de l'édition de *Diary of a Century* (1970) ».

La présentation proposera un descriptif de la base de données personnalisée à l'œuvre de Lartigue. On est ici dans le cas d'une interface regroupant un corpus monographique d'une grande cohérence : l'ensemble de la production est accessible (16 000 pages d'albums, dessins et peintures photographiés) ainsi que le corpus d'archives relatives à l'œuvre : les agendas et journaux intimes de 1911 à 1986 (scannés), la correspondance (9000 documents scannés) et la presse (articles de presse scannés). L'outil a d'abord cette fonction de regrouper toutes les données dans une même unité, le tout est scanné pour faciliter la lecture. Un grand volume de données peut être ainsi appréhendé rapidement par des visualisations mettant sur le même plan images et données historiques. En mettant ainsi à plat une très grande quantité de données diverses, l'interface numérique engage véritablement des problématiques de recherche. Elle serait en quelque sorte l'équivalent d'une grande table de travail sur laquelle sont posés des milliers de documents, à la différence près qu'ici, il est possible de les visionner en très grand nombre d'un coup, de changer leur emplacement, de les faire se confronter sans cesse, de les classer et reclasser à l'infini.

En prenant en compte des données extérieures provenant de recherches spécifiques (archives consultées dans des institutions, archives privées...), en permettant surtout de conserver et de retrouver l'information (qui est l'un des soucis majeurs dans la recherche), l'outil s'avère être bien plus qu'une banque de données mais enclenche directement des réflexions, qui elles-mêmes engagent le développement de la base informatique.

On est ici dans un cas de projet de recherche intimement lié au développement d'une interface numérique monographique. L'enrichissement est double : les informations saisies dans la base et les connexions ainsi mises en place génèrent des problématiques de recherche qui elles-mêmes nourrissent la base de données. Le va-et-vient entre les deux est incessant : la saisie et la lecture des données vont de pair.

Dans le cas de Lartigue, l'étude a surtout montré, en déployant de multiples affichages de milliers d'images album par album, que les albums, loin d'être de simples albums de famille

constitués au fil du temps, s'organise au sein d'un système complexe, de démontage et remontage, sous l'influence d'un autre artiste, Richard Avedon.

Références

Le Galliard, Martine, «De la gestion de collection à l'analyse historique : la base de données Lartigue », *Études photographiques*, décembre 2006, pp. 141-152.

Moore, Kevin, *The Invention of an Artist*, Princeton, Princeton Press University, 2004.

Rice, Shelley / John Szarkowski, *Le choix du bonheur de Jacques Henri Lartigue*, Paris, Éditions La Manufacture / AAJHL, collection « Donation », n°6, 1992.

Sayag, Alain / Quentin Bajac (dir.), *Lartigue, l'album d'une vie*, Centre Pompidou, Éditions du Seuil, 2003.

Szarkowski, John, « The Photographs of Jacques Henri Lartigue », *The Museum of Modern Art Bulletin*, vol XXX, n°1

Fleur Pfeifer (Leipzig)

La communication scientifique ou: l'essai d'une méta-analyse

Il est présenté une étude du discours scientifique français, mettant surtout l'accent thématique sur les méthodes appliquées. L'étude examine des propos tenus dans des publications scientifiques en langue française en prenant en considération 1) le Lexique non-terminologique et 2) Énonciation et pragmatique. A cette fin, une méta-analyse ainsi que l'analyse d'un corpus sont effectuées. L'étude vise à recueillir des données qui pourront être exploitées dans le contexte de l'apprentissage des langues étrangères.

L'importance primordiale du discours pour la science est précisée par Kretzenbacher (1997: 134) quand il dit qu'il n'y a « *von der wissenschaftlichen Hypothesenbildung über die Stadien der Forschung bis zur Kommunikation und Diskussion von Forschungsergebnissen keinen 'sprachfreien' Raum.* » A côté du processus de recherche visé ici, le domaine du « discours scientifique » comprend tout propos tenu dans le contexte scientifique. L'examen de celui-ci, à partir de différents points de vue, fait partie de la recherche du discours scientifique.

Actuellement, l'anglais est la *Lingua franca* pour la majorité de la communauté scientifique internationale. Ainsi, il n'est pas surprenant que, dans la recherche du discours scientifique, le sujet « *English for Academic Purposes* » occupe une place prépondérante. D'autre part, « Le français comme langue scientifique » est également un domaine de recherche actif. Cela est démontré par la synthèse de Rinck (2010) qui classe les présentes études relatives au discours scientifique, en fonction de leurs objets d'étude, en quatre catégories : a) Genres de textes, b) Structure des textes, c) Lexique et d) Énonciation et pragmatique. Pour l'étude présentée ici, les catégories Lexique et Énonciation et pragmatique occupent une position centrale. La catégorie Lexique porte sur le lexique non-terminologique sous forme d'unités phraséologiques (cf. Tutin 2007; Leggallois & Tutin 2013) et la catégorie Énonciation et pragmatique met au centre de l'intérêt les fonctions pragmatiques de propos individuels. Afin de pouvoir formuler une généralisation sur ces deux catégories, une méta-analyse, qui fait la synthèse des résultats des études empiriques déjà publiées (cf. sur la méta-analyse statistique, Bortz & Döring 2006), sera effectuée dans une première phase d'étude.

Jusqu'à présent, la recherche montre que les études publiées ne mentionnent, le plus souvent, qu'un choix des exemples linguistiques les plus importants. Afin de vérifier et compléter ces exemples linguistiques, la méta-analyse est élargie d'une analyse d'un corpus linguistique. En principe, l'examen porte sur la question de savoir quelles sont les unités

phraséologiques qui se trouvent dans les publications scientifiques en langue française, et ce dans quel contexte et avec quelle fonction. La base de données sera le corpus *Scientext* (2013) qui est librement accessible.

En résumé, la méta-analyse ainsi que l'analyse du corpus s'appuient sur des propos traités numériquement. En rapport avec le sujet du congrès et en particulier de la section Interfaces numériques, cette étude montre les questions que soulève l'analyse linguistique des « représentations numériques ». Il reste cependant à examiner dans quelle mesure les résultats de la méta-analyse ainsi que de l'analyse du corpus peuvent être utilisés dans le contexte « Français Langue Étrangère ». La question n'y est pas de savoir si, mais dans quelle mesure un bénéfice existe et dans quelle proportion il sera élargi par des possibilités numériques.

Bibliographie

- Bortz, Jürgen / Döring, Nicola (2006): *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg: Springer.
- Kretzenbacher, Heinz L. (1997): "Fachsprache als Wissenschaftssprache", in: Hoffmann, Lothar / Kalverkämper, Hartwig / Wiegand, Herbert Ernst / Galinski, Christian / Hüllen, Werner (eds.): *An International Handbook of Special-Language and Terminology Research* (= HSK 14.1). Berlin: Walter de Gruyter 133-142.
- Leggallois, Dominique / Tutin, Agnès (2013): "Présentation: Vers une extension du domaine de la phraséologie", in: *Langages* 189: 3-25.
- Rinck, Fanny (2010): "L'analyse linguistique des enjeux de connaissance dans le discours scientifique: Un état des lieux", in: *Revue d'anthropologie des connaissances* 4: 3, 427-450.
- Scientext (2013): *Corpus Scientext*. LIDILEM, LICoRN, LLS
<<http://scientext.msh-alpes.fr/scientext-site-en/spip.php?article1>> [30.01.14].
- Tutin, Agnès (2007): "Autour du lexique et de la phraséologie des écrits scientifiques", in: *Revue française de linguistique appliquée* 2: XII, 5-14.

Bibliographie

- Bortz, Jürgen / Döring, Nicola (2006): *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg: Springer.
- Kretzenbacher, Heinz L. (1997): "Fachsprache als Wissenschaftssprache", in: Hoffmann, Lothar / Kalverkämper, Hartwig / Wiegand, Herbert Ernst / Galinski, Christian / Hüllen, Werner (eds.): *An International Handbook of Special-Language and Terminology Research* (= HSK 14.1). Berlin: de Gruyter 133-142.
- Leggallois, Dominique / Tutin, Agnès (2013): "Présentation: Vers une extension du domaine de la phraséologie", in: *Langages* 189: 3-25.
- Rinck, Fanny (2010): "L'analyse linguistique des enjeux de connaissance dans le discours scientifique: Un état des lieux", in: *Revue d'anthropologie des connaissances* 4: 3, 427-450.
- Scientext (2013): *Corpus Scientext*. LIDILEM, LICoRN, LLS.
<http://scientext.msh-alpes.fr/scientext-site-en/spip.php?article1> [30.01.14].
- Tutin, Agnès (2007): "Autour du lexique et de la phraséologie des écrits scientifiques", in: *Revue française de linguistique appliquée* 2: XII, 5-14.

Lars Schneider (München), Annette Gerstenberg (Berlin) & Christof Schöch (Würzburg)

La plate-forme romanistik.de comme interface numérique

romanistik.de (<http://www.romanistik.de>) est la plate-forme en ligne centrale de la romanistique de langue allemande. Fondée en 1999 par Albrecht Buschmann et Andreas Gelz sous forme de liste de diffusion puis site internet, romanistik.de est devenue une plate-forme d'information et de communication interactive sous la modération d'une rédaction. Aujourd'hui la plate-forme romanistik.de présente la romanistique germanophone dans toute sa pluralité disciplinaire. Les personnes travaillant dans le domaine de la romanistique peuvent y diffuser des informations concernant leurs activités de recherche et d'enseignement : annonces de postes, appels à contribution, programmes de colloques d'une part, projets de recherche en cours, publications récemment parues et profils personnels d'autre part. romanistik.de permet en outre aux associations professionnelles et aux sociétés universitaires issues de la romanistique allemande de présenter leurs activités sur des sites web dédiés. Dès lors romanistik.de est une image de la diversité et de la dynamique actuelle de la romanistique de langue allemande. La présente contribution voudrait retracer l'évolution de la plate-forme au fil des ans et montrer en quoi elle est devenue une véritable "interface numérique".

En effet, romanistik.de fonctionne comme une interface plurielle à travers laquelle les chercheurs et chercheuses en romanistique peuvent diffuser leurs activités, faire valoir leurs intérêts et s'informer des travaux récents et des possibilités de collaboration futures. De telle manière, la plate-forme crée un lien non seulement entre les chercheurs et chercheuses, mais également entre les activités « analogues » et leur diffusion dans l'espace numérique ainsi qu'entre les institutions et les individus impliqués dans la romanistique.

Les formes concrètes que cette interface a prises - depuis les débuts jusqu'à la refonte la plus récente du site en février 2014 - montrent qu'une interface numérique est tout sauf neutre. En effet, chaque structure technologique - liste de diffusion par courriel ou fil RSS, site web statique ou système dynamique de gestion des contenus - tend à influencer la présentation des sujets et leur assimilation par la communauté des chercheurs et chercheuses et facilite ou au contraire empêche certaines formes d'interaction et d'appropriation. Tout en montrant les objectifs, l'évolution et l'état actuel de la plate-forme, la contribution invitera également les participants et participantes à discuter les perspectives d'avenir qui s'offrent à romanistik.de.

Christof Schöch (Würzburg)

L'histoire du genre policier à travers ses "topics"

L'histoire du roman policier français s'écrit, le plus souvent, à travers la succession de plusieurs formes dominantes du genre et de ses précurseurs : tout commence avec le roman-feuilleton populaire et le roman judiciaire (p.ex. Eugène Sue, Émile Gaboriau) puis le roman d'aventures et d'exploits (p.ex. Gaston Leroux, Maurice Leblanc). Vient ensuite le récit de détection classique (p.ex. Georges Simenon), suivi par le roman noir (p.ex. Léo Malet), le roman à suspense (p.ex. Boileau/Narcejac) et le néo-polar (p.ex. Jean-Patrick Manchette, Didier Daeninckx). Une telle histoire du genre a pu s'appuyer sur des efforts typologiques tels que ceux de Tzvetan Todorov (1971) et à été écrite entre autres par Ulrich Schulz-Buschhaus (1975) en Allemagne ou par Boileau-Narcejac (1975) en France.

La présente contribution voudrait jeter un regard complémentaire sur un aspect précis de l'histoire du roman policier, un regard porté par les thématiques dominantes à telle ou telle époque et chez tel ou tel auteur. Elle présentera les résultats de l'étude d'une collection de

romans policiers ainsi que de leurs précurseurs, publiés entre 1860 et 2000 environ. Cette collection est explorée à travers la procédure du "topic modeling" développé par David Blei (2011) en utilisant l'outil MALLET (McCallum 2002). Cette procédure découvre de manière non-supervisée des groupes de mots thématiquement liées dans une collection de textes. La contribution voudrait éprouver l'utilité de la technique du "topic modeling" pour l'histoire d'un genre littéraire comme le roman policier. Elle voudrait également s'interroger sur la question de savoir en quoi le topic modeling, en tant qu'outil intermédiaire entre les textes et le chercheur, donc en tant qu'interface numérique (Douehi 2012), promeut ou au contraire empêche certaines manières d'envisager le genre policier. Elle sera l'occasion, aussi, de s'interroger sur les apports du paradigme des "humanités numériques" et plus précisément du "algorithmic criticism" (Ramsay 2011) aux études de lettres.

Références

- Blei, David M. 2011. "Introduction to Probabilistic Topic Models." *Communication of the ACM*.
- Boileau-Narcejac. 1975. *Le roman policier*. Paris: PUF.
- Douehi, Milad. 2012. *Pour un humanisme numérique* (2011). Paris: publie.net.
- McCallum, Andrew. 2002 "MALLET: A Machine Learning for Language Toolkit." <http://mallet.cs.umass.edu>. University of Massachusetts Amherst.
- Ramsay, Stephen. 2011. *Reading Machines. Toward an Algorithmic Criticism*. Urbana: Univ. of Illinois Press.
- Schulz-Buschhaus, Ulrich. 1975. *Formen und Ideologien des Kriminalromans : ein gattungsgeschichtlicher Essay*. Frankfurt am Main: Humanitas-Verlag.
- Todorov, Tzvetan. 1971. „Typologie du roman policier“. *Poétique de la prose*. Paris: Seuil, 55-65.

Susanne Schütz / Marcus Pöckelmann (Halle) **IT-Werkzeuge zur Unterstützung elektronischer Edition am Beispiel eines französischen Textes aus dem 18. Jahrhundert**

Im hier vorgestellten Projekt wird die Textgenese am Beispiel der Lateinamerikadarstellung in der *Histoire des deux Indes* nachvollzogen und als elektronische Edition präsentiert. Ausgangspunkt des Textvergleichs sind die vier Fassungen, die nach TEI-Standard in xml basiscodiert sind. Ziele des Projekts sind der semiautomatische Vergleich der Textfassungen und eine anschauliche Präsentation der Textvarianten nach Vorgaben der Herausgeber. In dieser Machbarkeitsstudie soll es nicht darum gehen, die eine endgültige Edition des Textes herzustellen, sondern den Philologen die Möglichkeit zu geben, den Textvergleich je nach Erkenntnisinteresse zu spezifizieren und die Visualisierung der Ergebnisse an das Editionsziel anzupassen. Im Zentrum der Arbeit steht also die Entwicklung von Methoden, die die Erarbeitung von philologischen Editionen mit informatischen Werkzeugen unterstützen. Der Vortrag präsentiert den Forschungsansatz, das methodische Vorgehen, erste Ergebnisse und offene Fragen, die bei der interdisziplinären Entwicklung von Editions Werkzeugen auftreten.

Weitere Informationen:

<http://informatik.uni-halle.de/ti/forschung/ehumanities/sada>

Dirk Siepman / Christoph Bürgel (Osnabrück)

Das Corpus de Référence du français contemporain

Der Vortrag stellt ein neues Referenzkorpus des Französischen vor (*Corpus de référence du français contemporain*; CRFC), das sich in Planung und Struktur an die großen Referenzkorpora des britischen und amerikanischen Englisch (BNC/COCA) anlehnt. Dieses 220-Millionen-Wort-Korpus setzt neue Maßstäbe nicht nur in Bezug auf die Textsortenbreite, sondern auch hinsichtlich der Repräsentation gesprochener bzw. pseudo-gesprochener Sprache. Besondere Aufmerksamkeit verdient der große Umfang des rein sprechsprachlichen Korpus (50 Mio. Wörter), ein Bereich, in dem die Forschung bisher auf kleinere, häufig durch Interviews elizitierte Daten zurückgreifen musste (vgl. Debaisieux 2010, Gadet et al. 2012). Dieses speist sich zu ca. 50% aus Transkriptionen von Fernsehsendungen der Sender France 2, France 3 und France 5.

Dies wirft zwangsläufig die Frage nach der Vergleichbarkeit solcher Daten mit der Sprechsprache außerhalb der Medien auf. Wie eine von Davies (<http://corpus.byu.edu/coca/>) vorgenommene Kontrastierung zeigt, besteht eine recht enge Passung zwischen den Transkripten der von ihm verwendeten Talkshows u.ä. und Sprechsprache außerhalb des Mediums Fernsehen. Darüber hinaus ist zu bedenken, dass das Korpus zahlreiche Fernsehdokumentationen enthält, die Interaktionen zwischen Sprechern zeigt, die sich der Beobachtung durch die Kamera nicht mehr bewusst sind, was die Natürlichkeit der Daten weiter steigert. Ergänzend zur Strukturierung nach Medium ist ein thematischer Querschnitt durch sämtliche Lebensbereiche gelegt worden, in Anlehnung an die Konzeption des *Oxford English Corpus* und des *Longman/Lancaster English Language Corpus*. Es ist unmittelbar einsichtig, dass eine zusätzliche thematische Ordnung insbesondere für wortschatzbezogene Forschung von Vorteil ist (vgl. Summers 1991: 7).

Das CRFC ermöglicht erstmals eine Grammatikschreibung, die verschiedene Register des Französischen im Sinne einer Gebrauchsgrammatik (Biber 1999, Deulofeu/Debaisieux 2012) berücksichtigt und kontrastiert. In diesem Bereich haben Bürgel/Siepman (2014, i.V.) eine Neuanalyse des Subjonctif-Gebrauchs vorgenommen, die insbesondere sprechsprachliche Konstruktionen in den Blick nimmt. Siepman (i.V.) legt eine auf dem CRFC basierende Untersuchung zur Repräsentation sprechsprachlicher Lexis in Wörterbüchern vor, die zeigt, dass sowohl die gängigen einsprachigen als auch zweisprachigen Wörterbücher des Französischen Eigenschaften der sprechsprachlichen Lexis bisher nur rudimentär erfassen.

Bibliographie

Bürgel, C. & Siepman, D. 2014. "Korpusinduzierte Grammatikschreibung des Gegenwartsfranzösischen am Beispiel des Subjonctif". Erscheint in: Tinnefeld, T. (2014) *Grammatik gestern-heute-morgen. Gedenksymposium für Hartmut Kleineidam*. Saarbrücken.

Deulofeu, H.-J. & Debaisieux, J.-M. 2012. "Une tâche à accomplir pour la linguistique française du XXIe siècle : élaborer une grammaire des usages du français". *Langue française* 176/4: 27-46.

Siepman, D. (i.V.) *Dictionaries and spoken language: a corpus-based review of French dictionaries*.

Summers, Della. (1991) "Longman/Lancaster English Language Corpus: Criteria and Design." Unpublished ms.

Suzan van Dijk (La Haye)

Collaborations rendues possibles grâce à Internet

Si on veut comprendre la participation féminine dans le champ littéraire européen des siècles précédant le XXe, il y a une réelle nécessité – me semble-t-il – de faire collaborer de façon systématique des professionnels et des amateurs. C'est ce que nous sommes en train d'expérimenter à La Haye à notre Institut Huygens pour l'Histoire des Pays-Bas (Huygens ING). On est en train de mettre sur pied une collaboration incluant trois catégories de participants : (1) des chercheurs en histoire littéraire, (2) des spécialistes ICT et (3) des gens intéressés et suffisamment qualifiés qui sont à la fois les lecteurs/trices de nos travaux et le public potentiel des écrivaines que nous étudions.

L'institut a été « grant holder » d'une Action européenne COST intitulée « Women Writers in History » (IS0901, 2009-2013). Dans ce cadre, un important réseau de chercheurs européens a étudié (et continue d'étudier) la participation des femmes écrivains dans le monde littéraire (jusqu'au début du XXe siècle). Cela se passe à une large échelle transnationale, en insistant sur la façon dont ces écrits ont été reçus par les lecteurs contemporains. Dans ce cadre, les historiens littéraires collaborent avec des développeurs, qui mettent à leur disposition un outil permettant cette approche large: une base de données intitulée WomenWriters. Cette Action COST a résulté (entre autres) en deux projets à nouveau internationaux : *Travelling Texts 1790-1914*, financé par HERA (2013-2016) et COBWWWEB, financé par CLARIN-NL (2013-2014). Ces deux projets – entre lesquels il y a également un lien de collaboration – représentent bien les deux aspects : contenu (histoire littéraire des femmes) et structure (base de données).

La collaboration avec le troisième groupe est essentielle : il s'agit du lectorat potentiel de « nos » écrivaines, qui illustre par sa présence la pertinence de nos recherches. Nous avons commencé par initier un projet de numérisation de la correspondance d'une seule écrivaine (d'importance européenne : Belle de Zuylen/Isabelle de Charrière, 1740-1805). La collaboration entre le projet large, visant la quantité et utilisant une structure pour classer les diverses catégories de textes « féminins » et ce projet plus ponctuel, qui se poursuit au niveau du texte dans lequel on peut annoter les éléments significatifs, permettra de traiter la problématique de façon plus complète. On peut dès lors regarder dans les deux directions et adopter les deux perspectives, qui en fait se répondent : Comment se positionnaient les femmes auteures ? Et comment étaient-elles perçues ? Les outils électroniques y correspondent, puisque les annotations dans le texte peuvent être identiques aux catégories établies dans la base ; et le cas individuel mène à des interrogations qui visent l'auctorialité féminine en tant que telle. Il s'agit là d'une problématique que – jusqu'à nouvel ordre – on ne pourra laisser à l'ordinateur : nous avons besoin de l'œil et du jugement humains entre le niveau du texte où l'auteure individuelle s'est exprimée et celui du réseau extrêmement complexe du champ littéraire tel que représenté dans une base de données.

Mais les personnes se situant ainsi entre deux n'ont pas besoin d'être des spécialistes. Ce type de lecture « dirigée » – assez intensive, il est vrai – est typiquement une activité pour laquelle nous comptons profiter grandement de l'aide de nos bénévoles, qui se trouvant dans une structure préétablie ne se sentent pas « perdus » et ont la possibilité de réellement contribuer.

Références

Gilleir, Anke / Alicia Montoya / Suzan van Dijk, « Before NEWW (New approaches to European Women's Writing). Prolegomena to the Launching of an International Project », in *Tulsa Studies in Women's Literature* 27, 1 (2008), pp. 151-157.

van Dijk, Suzan, « La correspondance d'Isabelle de Charrière en ligne », *Cahiers de l'Association Internationale des Etudes Françaises*, 64 (2012), pp. 29-40.

van Dijk, Suzan / Montserrat Prats Lopez, « Participation as a way of creating new audiences? », contribution to *Final conference of COST Action IS0901 « Women Writers In History »: Female authorship in Europe: Networks and obstacles*. The Hague, June 2013.

David Wrisley (Beirut)

Lieu, Temps, Réseau : la modélisation numérique des « géographies littéraires » du moyen âge français

Aujourd'hui les interfaces numériques sont au cœur de la géographie et elles risquent de s'imposer de plus en plus dans le comportement informationnel de futurs chercheurs en sciences humaines. Mon intervention porte sur de nouvelles pratiques numériques pour analyser les textes littéraires.

En l'occurrence il s'agit d'une base de données composée de toponymes figurant dans plus de cent textes littéraires du moyen âge français. L'analyse des toponymes dépend de la structuration des données, aussi bien par la localisation dans l'espace géomatique que par l'élaboration d'autres détails relatifs à leur production (des métadonnées telles que genre, traits formels, date et région de composition, dialecte).

A présent la base de données comprend des textes franco-italiens, anglo-normands et bourguignons, permettant déjà la comparaison du phénomène de la « géographie littéraire » dans des sources provenant de trois régions différentes.

Au cours de mon intervention je compte non seulement souligner les difficultés méthodologiques dans la collecte de telles données, mais aussi donner quelques exemples de leur traitement par la visualisation sur les cartes numériques. Je voudrais également montrer quelques exemples d'analyse des données géographiques *hors carte*, appliquant les principes de l'analyse des réseaux afin d'explorer le toponyme en tant que topique dans un réseau symbolique.