

“

**Corpus de textes :
composer, mesurer, interpréter**

”

Table des matières

<i>Usage combiné de Tropes® et d'un outil statistique pour recontextualiser et interpréter des données d'occurrences : Principe et illustration</i> , S. Soetewey [et al.].....	1
<i>Discours de presse et construction des problèmes publics : l'apport d'un traitement numérique des données</i> , A. Arnoult et L. Jacquez.....	3
<i>De la constitution d'un large corpus au recadrage des outils d'observation. L'exemple de la littérature des « nouveaux experts en sécurité »</i> , K. Delimitsos.....	4
<i>Étude lexico-chronologique d'un gros corpus de presse : les révolutions arabes dans l'AFP</i> , N. Souillard [et al.].....	6
<i>Revue de presse écrite : les PCB en corpus. Des archives au traitement de textes sous format numérique</i> , E. Comby [et al.].....	9
<i>Le roman comme source pour les recherches en patrimoine architectural. L'exemple des ambiances des maisons traditionnelles kabyles</i> , A. Belakehal.....	10
<i>Retour de pêche. Le métier de pêcheur à travers le discours des professionnels français du lac Léman</i> , Y. Le lay [et al.].....	11
<i>Paroles d'acteurs : en quoi une analyse de discours peut-elle éclairer la gestion des espèces invasives le long des cours d'eau ?</i> M. Cottet [et al.].....	13
<i>Quel corpus pour l'étude des usages discursifs de l'analogie ?</i> R. Ben Yacoub.....	14
<i>Quel corpus pour étudier les néologismes de la langue de l'économie en Algérie ?</i> S. Lanseur.....	16
<i>Un corpus numérique d'arabe : bilan et perspectives</i> , C. Pinon.....	17

“

Corpus numériques et méthodes

”

Usage combiné de Tropes® et d'un outil statistique pour recontextualiser et interpréter des données d'occurrences : Principe et illustration

Soetewey Sabine*, Derobertmeasure Antoine**, Duroisin Natacha***

* Sciences de l'éducation, Université de Mons, Institut d'Administration scolaire, 18 Place du Parc, 7000 Mons, Belgique, sabine.soetewey@umons.ac.be

** Sciences de l'éducation, Université de Mons, Institut d'Administration scolaire, 18 Place du Parc, 7000 Mons, Belgique, antoine.derobertmeasure@umons.ac.be

*** Sciences de l'éducation, Université de Mons, Institut d'Administration scolaire, 18 Place du Parc, 7000 Mons, Belgique, natacha.duroisin@umons.ac.be

1. INTRODUCTION

Confrontée à un corpus de textes de taille conséquente, l'analyse de contenu (Bénel, Lejeune et Zhou, 2010) manuelle devient difficile voire irréaliste pour le chercheur. Le recours à un logiciel automatique semble alors être une solution providentielle. Toutefois, procédure automatisée ne signifie pas, pour autant, analyse automatique. D'une part, face au choix de plus en plus vaste de solutions informatisées (Boutigny 2005, Séror 2005), le chercheur peut être perdu lorsqu'il s'agit d'identifier celle adaptée à son objet de recherche. D'autre part, au vu de la quantité de données générées par ces outils et de leur caractère dispersé, le chercheur peut être mis en difficulté au moment de les organiser et d'en extraire une signification.

2. DU CAS ANALYSE A LA METHODOLOGIE

Impliqués dans une analyse des programmes de cours de sciences (Soetewey, Duroisin & Demeuse, 2010), les auteurs se sont posé la question de la cohérence entre l'orientation pédagogique officiellement annoncée et le discours effectivement tenu au cœur d'un corpus de 15 programmes et référentiels. En prenant appui sur les travaux de Derobertmeasure et Demeuse (2011), relatifs au choix d'un logiciel, et ceux de Mangez & Mangez (2008), relatifs à l'analyse du discours, nous proposons et appliquons à un exemple concret, une méthode permettant à la fois une objectivation du relevé et de l'organisation des données et une interprétation contextualisée. Le principe consiste à combiner l'usage d'un outil d'analyse automatisé, Tropes®, à l'analyse en composante principale (ACP). Cette approche considère chaque élément relevé par Tropes® comme une dimension décrivant le corpus. L'ACP explore l'espace à n dimensions (des n éléments relevés) afin d'identifier une structure « résumant au mieux » le corpus. Cette restructuration de l'espace dans un nombre restreint de dimensions permet de faciliter le travail d'organisation des données. Tropes® a, quant à lui, été choisi pour sa capacité à extraire et faire émerger un contenu spécifique du texte, sans formatage théorico-contextuel préalable, en l'occurrence les verbes, potentiels indicateurs de l'orientation pédagogique caractérisant les programmes de cours. Suite à l'analyse textuelle réalisée avec Tropes®, relevant 4368 occurrences de verbes, l'analyse en composantes principales (ACP) permet d'envisager le corpus comme un espace d'autant de dimensions que de verbes, espace dans lequel chaque document est situé. Dans cet espace, l'ACP va identifier quelques nouvelles dimensions qui décrivent une part importante de la position des différents éléments du corpus dans l'espace. Dans notre exemple, l'ACP permet d'isoler 3 composantes décrivant 66% de la diversité au sein du corpus. La corrélation de ces composantes avec les dimensions de départ, permet d'associer chaque composante à une série de verbes. C'est à partir de cette étape que l'on quitte l'objectivation et que le travail d'interprétation et de contextualisation peut débuter, en prenant appui sur la littérature. Dans le cas analysé, trois documents atypiques ont pu être caractérisés, chacun, par une composante particulière, alors que le reste des documents présente une relative homogénéité. De cette façon, un programme ayant un score élevé sur la première composante a pu être associé à une approche de l'enseignement centrée sur des actions de base (refaire, exercer, calculer, ...) et de régulation (vérifier, corriger, ...), le faisant apparaître

comme relativement éloigné du discours officiel.

3. CONCLUSIONS

Face au défi visant à donner du sens à des données nombreuses et dispersées, le recours à l'analyse en composante principale offre la possibilité d'une structuration rigoureuse, objective et reproductible. De plus, elle autorise et favorise la recontextualisation des données extraites, au sens propre, de leur contexte de départ. L'usage du logiciel Tropes® permet, lui, une analyse sans formatage préalable qui laisse émerger le contenu du texte. Il permet également un retour dans les documents de référence afin de valider ou infirmer les interprétations contextuelles réalisées sur la base des données.

REFERENCES BIBLIOGRAPHIQUES

- Bénel A, Lejeune C, & Zhou C. (2010). « Éloge de l'hétérogénéité des structures d'analyse de textes », Document Numérique, 13(2), 41-56.
- Boutigny E. (2005). « Vers un renouvellement de la démarche qualitative en sciences de gestion? », Revue management et avenir, 2 (4), 59-69.
- Derobertmeasure A., Demeuse M. (2011). « Utilisation conjointe de deux logiciels d'analyse de contenu dans le cadre de l'analyse de traces de réflexivité : éléments de comparaison ». In Blais, J.-G. et Gilles, J.-L. (dir.) (2011). « Évaluation des apprentissages et technologies de l'information et de la communication : Le futur est à notre portée », Les Presses de l'Université Laval : Québec, p 163-189.
- Mangez E., Mangez C. (2008), "Analyse sociologique des discours pédagogiques. Application au cas de la politique éducative en Belgique francophone", In: Frandji D., Vitale P. ed(s), "Actualités de Basil Bernstein", Presses Universitaires de Rennes: Rennes, p. 189-206.
- Seror J. (2005). "Computers and qualitative data analysis: Paper, pens, and highlighters vs. screen, mouse, and keyboard", Tesol Quarterly, 39(2), p 321-328.
- Soetewey S., Duroisin N., Demeuse M. (2012), "Le curriculum oublié", Revue Internationale d'Education de Sèvres, 56, p 123-133.

Discours de presse et construction des problèmes publics: l'apport d'un traitement numérique des données

Lise JACQUEZ*

Audrey ARNOULT*

* Sciences de l'Information et de la Communication, Laboratoire ELICO, lise.jacquez@gmail.com et audrey.arnoult@univ-lyon2.fr

Cette proposition de communication s'appuie sur deux travaux de doctorat¹ : les controverses médiatisées autour des expulsions de sans-papiers à la fin des années 2000 et la médiatisation des troubles liés à l'adolescence de 1995 à 2009. Ces objets d'étude - au premier abord différent - s'inscrivent dans une réflexion commune sur la médiatisation d'un problème socio-politique dont l'objectif est de comprendre comment ces sujets sont mis en discours dans l'espace public. Quels sont les arguments, les logiques politiques ou encore les connaissances qui structurent les débats publics ? Sur quelles problématiques les médias se focalisent-ils ? Pour répondre à ces questionnements, nous analysons un important corpus de presse (plus de 1 000 articles) dont l'ampleur nous a conduites à utiliser le logiciel Modalisa pour disposer d'un premier aperçu de la médiatisation de ces deux objets. Dans le cadre de cette journée d'étude, nous souhaitons proposer une réflexion sur les apports d'un traitement numérique de corpus de presse pour repérer les modalités de la configuration d'un problème public dans les discours.

Dans cette perspective, le choix de l'usage d'un logiciel renvoie à plusieurs logiques (volume du corpus, volonté de disposer d'un premier aperçu de la mise en visibilité d'un sujet, analyse de contenu thématique, etc.) mais s'inscrit nécessairement dans une réflexion plus théorique :

- Quels principes ont guidé l'élaboration de nos grilles d'enquête respectives autrement dit, comment concilier objectifs de la recherche et contenu de l'enquête ?
- Dans quelles mesures les catégories définies pour le codage permettent-elles de répondre à nos problématiques de recherche ?
- Quelles types de catégories privilégier (numérique, textuel, à réponse unique ou multiple) pour disposer ensuite de résultats exploitables ?

En outre, le recours à un logiciel de traitement de texte pose également des questions d'ordre plus pragmatique :

- Le codage constitue souvent le point de départ d'une analyse quantitative donnant lieu à un traitement statistique. Or, nous souhaitons montrer qu'il est possible d'y intégrer des éléments qualitatifs.
- Enfin, nous voulons souligner la nécessité d'articuler le traitement numérique d'un corpus avec d'autres types d'approche des discours.

Cette communication nous permettra en résumé de questionner les choix épistémologiques et méthodologiques auxquels est confronté l'analyste de discours qui recourt à un logiciel pour traiter un important corpus de presse.

¹ La thèse de Lise Jacquez est en cours ; celle d'Audrey Arnoult a été soutenue le 8 décembre 2011.

“

**Emergence discursive
des représentations sociales**

”

De la constitution d'un large corpus au recadrage des outils d'observation.

L'exemple de la littérature des « nouveaux experts en sécurité ».

Konstantinos (Costa) DELIMITSOS

Doctorant en Sociologie - Criminologue, Univ. de Grenoble

ATER, Université de Lorraine, 2L2S-LASURES

Mail : konstantinos.delimitsos@univ-lorraine.fr

Résumé :

Ces dernières années se cristallise en France un mouvement prônant la constitution de la criminologie en tant que discipline universitaire spécifique. En tête de ce mouvement s'affichent des personnes à trajectoires personnelles et professionnelles bien différentes qui depuis 15 ans se sont érigées en experts en matière de sécurité. Aspirant à se positionner dans l'arène publique comme spécialistes, voire comme savants, ces acteurs déploient une stratégie d'acquisition de visibilité nationale et de légitimité professionnelle qui s'appuie fortement sur une riche production intellectuelle. Dans le cadre d'une thèse en sociologie, et par le biais d'une analyse textuelle, nous nous intéressons à la rhétorique, les formes d'argumentation et les représentations charriées par ces « nouveaux experts en sécurité ». Tirée de ce travail de recherche, cette communication se penchera sur deux points particuliers ; la construction du terrain et la mobilisation des outils informatiques permettant d'envisager une analyse de discours à grande échelle.

En ce qui concerne la constitution du corpus, nous nous proposons d'en retracer le chemin tout en mettant à l'exergue les difficultés rencontrées, touchant notamment aux questions d'accessibilité, de représentativité et d'homogénéité du terrain. Évoluant au sein et à la lisière des différents champs en y assurant différentes fonctions, les « nouveaux experts » affichent une littérature marquée par une très grande diversité et une accessibilité partielle. Pour y faire face, nous avons choisi de privilégier la piste des synergies bibliographiques réitérées autour des publications commercialisés traitant de la délinquance en France, piste amenant à la constitution d'un corpus de 24 ouvrages « grand public ». Or, un travail d'analyse de contenu et de discours sur un corpus aussi important ne serait pas aisément envisageable en l'absence des outils informatiques spécialisés, en l'occurrence des logiciels de reconnaissance optique des caractères permettant sa numérisation ainsi que des logiciels d'analyse lexicométrique. Il n'en reste pas moins que lors de nos recherches, nous nous sommes heurtés à des limites que présentent à leur tour ces logiciels. Différentes stratégies ont dû être développées alors, visant à remédier la fiabilité certes grande mais non absolue qui est l'attribut des outils informatiques ouvrant l'accès aujourd'hui à des corpus bien plus grands qu'auparavant...

La littérature des « nouveaux experts en sécurité »

- 1 - BAUER A. (2010), *A la recherche de la criminologie. Une enquête*, Paris, Éditions CNRS.
- 2 - BAUER A. (2011), *Criminologie plurielle. Une introduction générale à la criminologie*, Paris, PUF.
- 3 - BAUER A. et FREYNET Fr. (2008), *Vidéosurveillance et vidéoprotection*, Paris, PUF, QSJ ?.
- 4 - BAUER A. et FREYNET Fr. (2009), *Les études de sûreté et de sécurité publique*, Paris, PUF coll. QSJ ?.
- 5 - BAUER A. et PEREZ E. (2009), *Les 100 mots de la police et du crime*, Paris, PUF, coll. QSJ ?.
- 6 - BAUER A. et RAUFER X. (1998), *Violences et insécurité urbaines*, Paris, PUF, coll. QSJ ?.
- 7 - BAUER A. et RAUFER X. (2009), *La face noire de la mondialisation*, Paris, Éditions CNRS.
- 8 - BAUER A., RAUFER X. et SOULLEZ Chr., (2006), *Violences et insécurité urbaines*, Paris, PUF, coll. QSJ ?.
- 9 - BAUER A., RIZK C. et. SOULLEZ Chr. (2010), *Statistiques criminelles et enquêtes de victimation*, Paris, PUF, coll. QSJ ?.
- 10 - BAUER A. et. SOULLEZ Chr. (2007), *Violences et insécurité urbaines*, Paris, PUF, QSJ ?.
- 11 - BAUER A. et. SOULLEZ Chr. (2009), *Les fichiers de police et de gendarmerie*, Paris, PUF, coll. QSJ ?.
- 12 - BAUER A. et. SOULLEZ Chr. (2011), *Les politiques publiques de sécurité*, Paris, PUF, coll. QSJ ?.
- 13 - BAUER A. et VENTRE A.-M. (2001), *Les Polices en France : sécurité publique et opérateurs privés. Le débat, les chiffres*, Paris, PUF, coll. QSJ ?.
- 14 - BOUSQUET R. (1998), *Insécurité. Nouveaux risques. Les quartiers de tous les dangers*, Paris, L'Harmattan. Préface d'A. Bauer et d'A.-M. Ventre.
- 15 - BOUSQUET R. (1999), *Insécurité. Nouveaux enjeux. L'expertise et les propositions policières*, Paris, L'Harmattan, préface d'A.-M. Ventre.
- 16 - BOUSQUET R. (2002), *Insécurité. Sortir de l'impasse*, Paris, LPM.
- 17 - BOUSQUET R. et LENOIR E. (2009), *La prévention de la délinquance*, Paris, PUF.
- 18 - RAUFER X. (2002), *L'explosion criminelle*, Paris, Valmonde et cie.
- 19 - RAUFER X. et QUERE St. (2002), *Le crime organisé*, Paris, PUF, coll. QSJ ?.
- 20 - RUDOLPH L. et SOULLEZ Chr. (2000), *La police en France*, Toulouse, Milan.
- 21 - RUDOLPH L. et SOULLEZ Chr. (2002), *Insécurité. La vérité*, Paris, J.-C. Lattès.
- 22 - RUDOLPH L. et SOULLEZ Chr. (2007), *Les stratégies de la sécurité 2002-2007. Avec 150 propositions pour aller plus loin*, Paris, PUF.
- 23 - SOULLEZ Chr. (2006), *Lexique de la sécurité*, Paris, PUF.
- 24 - SOULLEZ Chr. (2006), *Les violences urbaines*, Toulouse, Milan.

Étude lexico-chronologique d'un gros corpus de presse: les révolutions arabes dans l'AFP

Natacha Souillard*, Pierre Ratinaud**, Pascal Marchand***

* Sciences Politiques, Institut d'Études Politiques de Toulouse, Université de Toulouse , nat.souillard@gmail.com

** Sciences de l'Éducation, LERASS, Université de Toulouse, ratinaud@univ-tlse2.fr

*** Sciences de l'Information et de la Communication, LERASS, Université de Toulouse, pascal.marchand@iut-tlse3.fr

Les dépêches de l'AFP (Agence France Presse) représentent un terrain discursif pertinent pour étudier l'émergence et l'évolution d'objets médiatiques. La disponibilité de ces dépêches dans la base de données Factiva, associée à la possibilité de les extraire à partir de requêtes et à l'importation automatique des fichiers extraits dans certains logiciels de lexicométrie (TXM et Iramuteq) permet d'envisager l'étude exhaustive de la chronologie de la construction de ces objets. Nous appliquerons ce processus à l'étude des événements liés aux « révolutions arabes » sur une période de 9 mois.

L'ÉLABORATION DU CORPUS : CHOIX DES MOTS CLESET INDICATEURS TEMPORELS

Le corpus sur lequel repose cette étude est exclusivement composé de dépêches AFP, du mois de décembre 2010 au mois d'août 2011 sélectionnées à partir des mots clés Tunisie, Egypte, Libye et Syrie.

Ce corpus regroupe 21 055 dépêches et a été découpé en 219 364 segments de textes (contenant en moyenne 36,2 mots). Il comprend 7 939 978 occurrences correspondant à 70 246 formes et 48 718 lemmes (avec 21 405 hapax avant lemmatisation, soit 30,47% des formes).

METHODES DE CLASSIFICATION DES DONNEES TEXTUELLES

Nous avons soumis ce corpus à une classification hiérarchique descendante (CDH) avec l'algorithme de Reinert (1983) disponible dans le logiciel Iramuteq (Ratinaud & Déjean, 2009 ; Ratinaud & Marchand, 2012), en conservant 7840 formes actives. L'analyse retenue présente 36 classes terminales. Nous projetons ensuite, sur un axe chronologique, le χ^2 de liaison des dates (codées en jours de publication des dépêches) aux classes (cf. Figure 1).

Nous pouvons alors procéder à une lecture chronologique des mondes lexicaux extraits par la CDH qui permet notamment de mettre en évidence :

- La temporalité attendue des événements tels qu'ils se sont déroulés et succédés et donc ont été traités par les médias ;
- L'adaptation progressive (ruptures et continuités) du lexique au traitement de faits de nature différente, pourtant rassemblés sous le même référent « révolutions (ou printemps) arabe(s) »

Des exemples tirés de l'analyse illustreront ces observations.

ILLUSTRATION

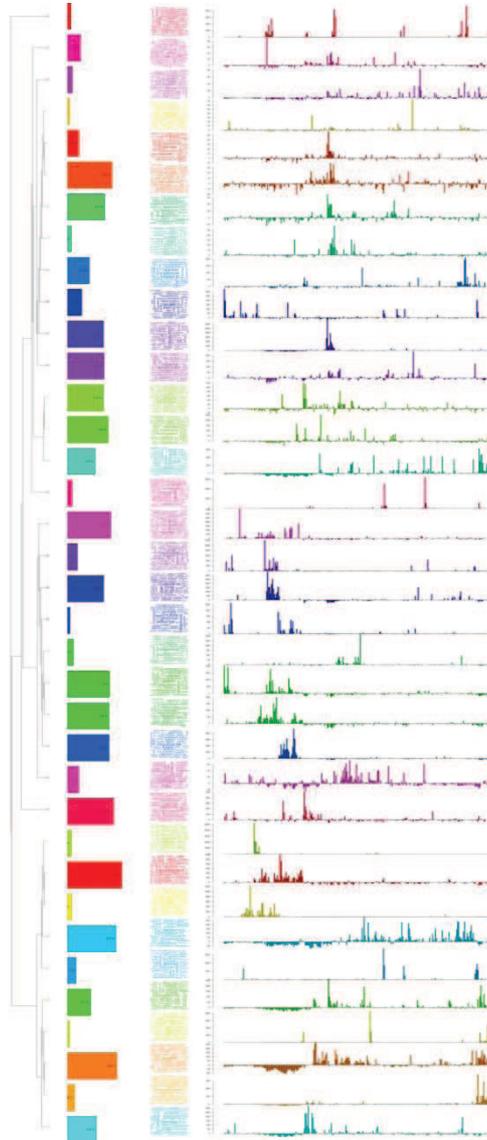


Figure 1. Dendrogramme, taille, nuage de mots des profils et projection chronologique des classes (taille réelle : 2000x6000)

REFERENCES BIBLIOGRAPHIQUES

- Ratinaud, P., & Déjean, S. (2009). « IRaMuTeQ : implémentation de la méthode ALCESTE d'analyse de texte dans un logiciel libre ». Présenté à *Modélisation Appliquée aux Sciences Humaines et Sociales* (MASHS2009), Toulouse, France
- Ratinaud, P., & Marchand, P. (2012). « Application de la méthode ALCESTE à de « gros » corpus et stabilité des « mondes lexicaux » : analyse du « CableGate » avec IRaMuTeQ ». In *Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles*. 835-844
- Reinert, M. (1983) : « Une méthode de classification descendante hiérarchique : application à l'analyse lexicale par contexte », *Les Cahiers de l'analyse des données*, Vol. VIII, n° 2 : 187-198.

Revue de presse écrite : les PCB en corpus.

Des archives au traitement de textes sous format numérique

Emeline Comby, Yves-François Le Lay, Hervé Piégay*

* Géographie, Université de Lyon, UMR 5600 Environnement Ville Société, mail du correspondant : emeline.comby@ens-lyon.fr

En 2005, la pollution du Rhône par les PCB est redécouverte, après une première crise au milieu des années 1980 et un oubli dans la décennie 1990. Dès septembre 2005, cet événement reçoit un écho important dans la presse quotidienne régionale lyonnaise. Néanmoins, la diffusion médiatique de cette information n'est pas uniforme sur ce linéaire fluvial français qui est pourtant entièrement pollué : ces faits ne sont jugés dignes d'intérêt dans la presse nationale et celles du Rhône moyen et du Bas Rhône qu'au cours de l'année 2007, quand les interdictions de pêche et les analyses scientifiques se font plus nombreuses. Cette contribution vise a) sur le plan méthodologique à présenter les avantages et les inconvénients de la numérisation d'un corpus de quatre quotidiens et à montrer la complémentarité de différentes offres logicielles gratuites dans une approche mêlant quantitatif et qualitatif ; et b) à aborder points communs et contrastes entre quatre discours médiatiques sur une même crise socio-environnementale.

1. DEUX TYPES DE RECUEIL DE LA DONNÉE : MATÉRIEL ET MÉTHODES DE TRAITEMENT DE CORPUS NUMÉRIQUES

Deux types de corpus sont créés selon l'aire de diffusion des titres (presse nationale ou presse régionale), leur disponibilité dans des bases de données payantes et la volonté de dépouiller le journal tel qu'il est lu par les lecteurs. Pour la presse quotidienne régionale, le recueil des articles sous format papier de 2005 à 2012 s'est fait en trois lieux : au Centre de documentation contemporaine de l'IEP de Lyon pour Le Progrès (édition Lyon, Villeurbanne, Caluire), aux Archives départementales de Valence pour Le Dauphiné Libéré (éditions Drôme – Ardèche) et à la médiathèque Ceccano d'Avignon pour La Provence (édition Grand Avignon). Les articles sont photographiés et renseignés dans un tableau, avant d'être transformés via un logiciel de reconnaissance de caractères. Pour la presse quotidienne nationale (Le Monde), le dépouillement s'effectue via la base de données payante Europresse, par la succession de deux requêtes de 1987 à 2012 « Rhône » et « PCB ». Cette méthode demande un temps de sélection pour éviter les doublons, mais permet d'éviter les oublis fréquents lors de l'utilisation d'une seule requête. 150 articles sont traités via une analyse de contenu dans R et des analyses de données textuelles avec TXM et IraMuTeQ. L'accent est mis sur une offre gratuite qui connaît actuellement de nombreux développements.

2. QUATRE VISAGES DE LA CRISE EN MIROIR : QUELQUES RESULTATS ET ELEMENTS DE DISCUSSION

Une partition du corpus par année permet de montrer une évolution des discours, notamment via des progressions et des analyses factorielles de correspondances. Les acteurs cités changent quand la crise prend une ampleur nationale en 2007, en affirmant le rôle de l'Etat à travers la figure ministérielle ou préfectorale. Néanmoins, la diffusion spatiale de la crise semble aussi entraîner une plus forte homogénéité des discours, à l'heure où les informations et les interlocuteurs sont plus facilement identifiables. Ainsi, la prise en compte du contexte (notamment politique, temporel ou spatial) du corpus est nécessaire pour explorer des corrélations entre différents éléments factuels (via la prise en compte des métadonnées) et interpréter les résultats.

LE ROMAN COMME SOURCE POUR LES RECHERCHES EN PATRIMOINE ARCHITECTURAL. L'EXEMPLE DES AMBIANCES DES MAISONS TRADITIONNELLES KABYLES

Nadia ZIDELMAL et Azeddine BELAKEHAL
Architectes, Département d'architecture
Faculté des Sciences et de Technologie
Université de Biskra Algérie.
Email: nadiremas@gmail.com

L'habitat traditionnel est un héritage qu'il est nécessaire de transmettre aux générations futures pour ses nombreuses qualités révélatrices d'un état de société avec ses racines, ses repères et ses valeurs propres. Il permet de mieux cerner le mode de vie qu'avaient adopté nos ancêtres pour répondre à leurs besoins. Ce patrimoine a fait l'objet de plusieurs études de recherches et d'opérations de restaurations à travers le monde. Ces dernières tentent de le comprendre, à travers son style, son histoire ses éléments architectoniques, ses systèmes constructifs, le fonctionnement de ses édifices, également son aspect esthétique, formel et décoratif. Mais rares sont celles qui s'intéressent au vécu de ces espaces et aux différentes sensations qui s'en dégagent et qui sont perçues par ses usagers.

La présente recherche s'inscrit dans cette perspective en proposant une nouvelle façon d'aborder le patrimoine architectural ; celle qui prend en considération, en plus des aspects typiquement objectifs, des aspects multi- sensoriels tels que le toucher, l'audition, le goût, la vue, également la lumière et la thermique. Ces informations participent aux qualités sensorielles des lieux et leur incidence sur le bien-être, les sentiments de confort, voire les comportements. Ces notions sont complexes car elles ne dépendent pas uniquement des propriétés chimiques ou physiques d'un environnement, mais relèvent également du processus de perceptions et d'évaluations.

Ce rapport aux sens est recherché au moyen de la notion d'ambiance, et à travers un patrimoine architectural algérien à savoir la maison traditionnelle en Kabylie. L'objectif recherché, nous met devant la difficulté de réactiver des ambiances dans un espace qui n'est pas habité ou s'il l'est, est occupé d'une manière différente de celle originale. En face de cette situation, le recours à des sources textuelles, s'avère incontournable puisqu'elles constituent les seuls témoignages préservés. Etant cela, l'intérêt s'est porté particulièrement sur des romans écrits par des auteurs de la région et ayant connu ces maisons bien avant qu'elles subissent les effets des mutations socio-économiques et des transformations urbaines et architecturales. Ces romans s'avèrent autant intéressants que crédibles puisqu'ils relatent des faits déroulés dans les espaces en question et rapportent fidèlement le vécu et les perceptions qu'avaient les usagers kabyles pour leurs maisons. Il en advenait que la mesure, au moyen du texte, permettrait d'aboutir aux objectifs de cette recherche à savoir cerner le vécu sensoriel d'antan dans ce patrimoine architectural de la Kabylie.

L'analyse de contenu effectuée pour quatre romans a mis en évidence plusieurs types d'ambiances caractérisant la maison traditionnelle kabyle, ainsi que plusieurs composantes spatiales sources d'ambiances. Ces dernières constitueraient d'authentiques référents ambiantaux autant pour les opérations de sauvegarde de maisons traditionnelles que pour la conception de nouveaux projets d'habitat en Kabylie.

“

Etude en corpus clos

”

Retour de pêche. Le métier de pêcheur à travers le discours des professionnels français du lac Léman

Yves-François Le Lay*, Serge Heiden**, Luc Merchez* et Bénédicte Pincemin**

* Géographie, Université de Lyon, UMR 5600 EVS, mail du correspondant : yves-francois.le-lay@ens-lyon.fr

** Linguistique, Université de Lyon, UMR 5191 ICAR

La pêche en eau douce véhicule quelques images stéréotypées, parmi lesquelles celle d'un homme – âgé, barbu et taiseux – qui pose ses filets puis les lève, seul sur son embarcation. Les pratiques ont pourtant enregistré de réelles mutations depuis la Seconde Guerre mondiale, même si de vifs contrastes sont observés à la surface du globe (Welcomme et al., 2010). En novembre 2011, un groupe de géographes a profité d'un stage de Master pour mettre à l'épreuve les poncifs. Ils ont réalisé une enquête auprès des 46 pêcheurs professionnels recensés au registre de l'Association Agréée Interdépartementale des Pêcheurs Professionnels des Lacs Alpains (AAIPPLA). Si l'étude a reposé sur la complémentarité de l'analyse de données spatiales, d'un questionnaire et d'une campagne d'entretiens semi-directifs, seul le corpus constitué à partir de ces derniers sera présenté dans cette communication dont l'objectif est d'identifier, à travers les mots employés par les pêcheurs eux-mêmes, quels sont les principaux enjeux actuels de leur activité et quelles valeurs ils donnent aux évolutions récentes.

1. MATERIEL ET METHODE : ANALYSE TEXTOMETRIQUE D'UN CORPUS ORAL

Les enquêteurs ont d'abord réalisé une grille d'entretien commune autour de cinq thèmes majeurs : (a) la formation et le parcours de vie des pêcheurs, (b) leur équipement et les techniques de pêche employées, (c) la gestion des ressources halieutiques, (d) la patrimonialisation de l'activité, et (e) les jeux d'acteurs et les éventuels tensions et conflits. 38 des 46 pêcheurs professionnels de la rive gauche du lac Léman ont accepté de répondre à l'enquête. Les entretiens ont été enregistrés au moyen d'un enregistreur numérique puis ont été intégralement retranscrits avec Transcriber. Le corpus a été importé dans la plateforme textométrique open-source TXM qui propose des outils divers : lexique alphabétique et hiérarchique, index, concordancier, spécificités lexicales, progression, analyses factorielles et clustering hiérarchique (Heiden et al., 2010).

2. QUELQUES RESULTATS ET ELEMENTS DE DISCUSSION : QUELLE DURABILITE DE LA PECHE ?

Le discours tenu à l'égard du métier varie selon les pêcheurs et notamment en fonction de leur âge. Sur la figure 1, le plan factoriel F1 X F2 procède d'une analyse factorielle des correspondances (AFC) qui a été réalisée sur le corpus d'entretiens partitionné selon cinq classes d'âges (Lebart et al., 1998). Sur l'axe F1, la classe des interviewés les plus âgés s'oppose à celles des plus jeunes. De fait, les professionnels se positionnent diversement face mutations contemporaines. Certes la pêche est une activité traditionnelle, mais elle s'avère désormais tributaire des flux internationalisés de personnes, de biens et d'informations. A l'amont de la filière, les filets ne sont plus fabriqués dans la région, mais sont importés d'Asie ou d'Italie. L'activité quotidienne des pêcheurs est régie par un règlement franco-suisse et peut être localement perturbée par les navettes qu'effectue la compagnie de navigation entre la Suisse et la

France. A l'aval de la filière, lorsqu'ils cherchent à écouler leur production de perches, les pêcheurs du bassin lémanique se trouvent en concurrence avec ceux travaillant en Pologne ou en Russie. Le vocabulaire employé souligne combien les professionnels ont modernisé leur équipement. La barque en plastique s'est substituée à la barque en bois. Des vire-filets facilitent et accélèrent la remontée des filets. Autrefois en coton, ces derniers sont désormais en nylon. Surtout, seul un petit nombre de pêcheurs se contente aujourd'hui de pêcher. Ils accomplissent de nouvelles tâches liées à la transformation du poisson qu'ils doivent fileter, désarêter, voire fumer et conditionner. Mais il demeure quelques résistances à ces évolutions. Le métier repose encore sur des savoirs et des techniques traditionnels. Ainsi, quelques professionnels prennent encore plaisir à pratiquer régulièrement la pêche à la senne. De plus, malgré l'abondance de poissons en 2011, de nombreux entretiens soulignent les contraintes, liées notamment à l'importance des horaires de travail et à l'individualisme des pêcheurs : le collègue est le premier concurrent, si bien que leur association peine à s'organiser. Il reste que le sentiment d'être privilégié imprègne leur discours.

ILLUSTRATION

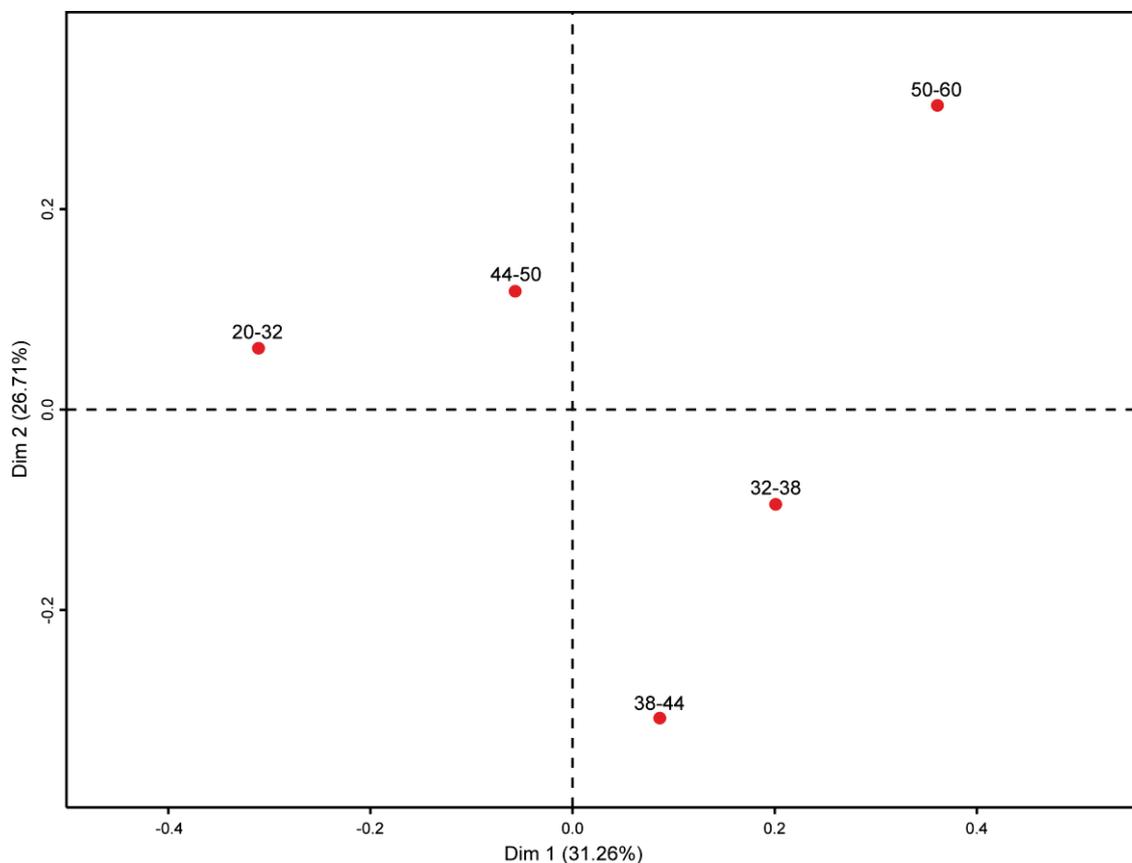


Figure 1. Des discours distincts en fonction de l'âge des pêcheurs évoqués

RÉFÉRENCES BIBLIOGRAPHIQUES

- Lebart L., Salem A., Berry L. (1998), *Exploring textual data*, Kluwert Academic Publishers, Dordrecht, 245 p.
- Welcomme R.L., Cowx I.G., Coates D., Béné C., Funge-Smith S., Halls A., Lorenszen K. (2010), « Inland capture fisheries », *Philosophical Transactions of The Royal Society B*, n° 365, p. 2881-2896.
- Heiden S., Magué J.-P., Pincemin B. (2010), "TXM: une plateforme logicielle open-source pour la textométrie - conception et développement", in I. Sergio Bolasco C. (éd.), *Proceedings of 10th International Conference on the Statistical Analysis of Textual Data - JADT 2010*, Edizioni Universitarie di Lettere Economia Diritto, Roma, tome 2, p. 1021-1032.

Paroles d'acteurs : en quoi une analyse de discours peut-elle éclairer la gestion des espèces invasives le long des cours d'eau ?

Cottet Marylise*, Piola Florence**, Rivière-Honegger Anne***

* Géographie, UMR 5600 EVS, Université de Lyon, ENS de Lyon, 15 parvis René Descartes, 69007 LYON, marylise.cottet@ens-lyon.fr

** Ecologie, UMR 5023 LEHNA, Université de Lyon, Université Lyon 1, Bâtiment Forel, 2ème étage, 43, Boulevard du 11 novembre 1918 69622 Villeurbanne Cedex, florence.piola@univ-lyon1.fr

*** Géographie, UMR 5600 EVS, Université de Lyon, ENS de Lyon, 15 parvis René Descartes, 69007 LYON, honegger@club-internet.fr

La gestion des cours d'eau est en partie basée sur les connaissances scientifiques acquises au sujet du fonctionnement des écosystèmes. Cependant, l'action environnementale est le plus souvent amenée à s'abstraire de ce cadre objectif et à s'appuyer sur les représentations des acteurs impliqués dans la gestion. D'une part, les connaissances scientifiques ne permettent pas d'appréhender toute la complexité des processus qui animent le vivant, produisant par conséquent des incertitudes pour définir les meilleures modalités d'action. D'autre part, la préservation des écosystèmes ne peut être exhaustive et doit reposer sur des choix concertés. Or, ces choix dépendent des représentations des acteurs de l'eau vis-à-vis des milieux. L'analyse de ces représentations constitue donc un axe de recherche incontournable, aux côtés des recherches conduites par les SVT, pour optimiser la gestion des cours d'eau. Elle permet du moins d'éclairer les choix qui sont opérés. L'analyse des discours produits par les acteurs de l'eau est un moyen efficace pour appréhender les représentations qui sont associées à une problématique de gestion.

Les questions relatives à la gestion de la renouée du Japon constituent un cas d'étude particulièrement éclairant. Cette plante est une espèce invasive très présente sur les berges de cours d'eau. Elle fait l'objet d'actions de lutte intensives par les gestionnaires, dont l'efficacité reste à l'heure actuelle limitée. Les actions menées (fauche notamment) contribuent même à accentuer la dispersion de la plante si peu de précautions sont prises. Les incertitudes relatives à la gestion de cette espèce invasive dans les sphères gestionnaires – pourquoi intervenir et comment agir efficacement ? – demandent d'acquérir une meilleure connaissance des représentations des acteurs de l'eau envers cette plante : quelles sont les raisons invoquées pour justifier d'une action à son encontre et pourquoi recourir plus particulièrement à tel ou tel mode de gestion ?

Afin de répondre à ces questions, un corpus de textes a été créé à partir de la documentation produite par les acteurs de l'eau impliqués dans la gestion du fleuve Rhône (guides techniques, plaquettes de communication, articles de journaux spécialisés...). Ces acteurs interviennent à différentes échelles territoriales (locale, départementale, régionale et de bassin). L'analyse du corpus couple une approche qualitative et quantitative (analyse statistique de données textuelles). Des éléments contextuels de production des discours ont été pris en compte pour les traitements. Ils se rapportent en particulier à des facteurs géographiques (étude des contrastes amont/ aval) et à des facteurs humains (étude des contrastes entre les différents types d'acteurs).

Les résultats montrent que les représentations associées à la renouée du Japon se rapportent avant tout aux actions de lutte. On observe un déséquilibre géographique de la production de discours (amont-aval). Les résultats confirment par ailleurs l'hypothèse d'un contraste des représentations entre les différents acteurs considérés. Des postures spécifiques peuvent être mises en évidence en fonction du profil des acteurs (échelle territoriale d'action et vocation de la structure gestionnaire). Cette variabilité des discours et des représentations sera discutée du point de vue des stratégies à développer pour optimiser la gestion de la renouée.

“

Corpus et linguistique : une évidence ?

”

QUEL CORPUS POUR L'ÉTUDE DES USAGES DISCURSIFS DE L'ANALOGIE ?

Rim Ben Yacoub

Linguistique générale, Institut Supérieur des Langues de Tunis, 91 rue Ibn Douraid, 2063, Nouvelle Médina, Tunisie, rimbenyacoub@yahoo.fr

Résumé :

La langue en usage élabore des stratégies et des modes de fonctionnement qui donnent, dans une ancienne définition de la rhétorique, une large place à l'écart conceptuel. L'écart entraîne des mécanismes saisis en dehors de la langue pour contribuer à plus d'expressivité ou à plus d'adéquation avec les intentions et / ou les discours mis en jeu. Par conséquent, la langue intègre les traces du parcours du sens. Un énoncé comme *Jean est un lion* par exemple, met en présence un rapport de *ressemblance* entre *Jean*, être humain et le terme *lion*, pris non dans son sens purement linguistique mais par rapport à un sens contextuel relatif à une circonstance discursive particulière dans laquelle *l'un des qualités de Jean* est mise en relief, *la force* par exemple. Ce rapport de similitude entre les deux pôles de la relation mise en place dans cet énoncé est un rapport *analogique*. Ces constatations nous ont amenée à réfléchir à la question de l'analogie comme descriptif d'une dynamique du parcours du sens entre discours et langue. Autrement dit, l'étude que nous avons proposée dans notre travail de thèse s'est faite dans le point de rencontre entre langue et discours. Expliquons-nous : nous pensons que l'analogie est un processus qui prend forme et se déploie à travers et par la langue, autrement dit, le lexique, mais s'actualise et reçoit une interprétation effective dans le discours c'est-à-dire à travers le contexte dans lequel elle est émise, et par rapport aux intentions, aux représentations et à l'ontologie relatives aux sujets parlants. Par conséquent, notre problématique était la suivante : Etant donné que l'analogie intègre dans son fonctionnement aussi bien le linguistique que l'extralinguistique, nous pourrions supposer qu'elle pourrait avoir un fonctionnement discursif susceptible de décrire une dynamique de sens entre langue et discours ; ce qui dépendra essentiellement du degré d'écart en question. Pour résoudre une telle problématique, il fallait recourir à un corpus d'exemples d'analogie circulant dans l'usage de la langue. Néanmoins, la dimension discursive que nous cherchons à relever nous a amené à analyser les différentes séquences collectées en les ancrant dans leur contexte extralinguistique. Ce faisant, nous nous sommes heurté à une question primordiale dans le cadre de l'élaboration de notre corpus concernant le type de corpus à choisir, l'approche à emprunter pour l'analyser et les conséquences qui en résultent.

En effet, parler des usages discursifs de l'analogie implique nécessairement une réflexion sur le type de corpus à choisir en cours d'analyse. Il s'agit d'un corpus qui doit être puisé dans des productions langagières usitées dans la vie quotidienne. Cette dimension pragmatique de l'analyse fait du corpus un corpus que nous pouvons qualifier de « pragmatique », représentant une réalité qui préexiste à son élaboration mais qui n'est cernable que par et à travers lui d'où le caractère empirique du corpus en question.

Par ailleurs, le caractère contrastif de l'étude de l'analogie que nous proposons dans notre thèse impose l'extension du corpus du cadre étroit d'une seule et unique langue vers un cadre plus ouvert et promettant qui est celui d'un va-et-vient permanent entre deux langues différentes du point de vue syntaxique, morphologique et culturel qui sont le français et l'arabe standard et dialectal. Ceci nous a permis de relever un paramètre déterminant qui est celui des variantes. Ces variantes nous ont permis de mettre en relief la spécificité de l'approche que nous proposons à l'étude analogique qui est de nature strictement pragmatique, autrement dit, d'une étude qui se fait à la lumière de l'ancrage des emplois analogiques au sein du contexte socioculturel dans lequel ils se déploient et par rapport à des considérations ontologiques relatives au sujet parlant. Par conséquent, ces variantes ont le mérite d'étendre le corpus vers d'autres horizons dans lesquels se déploient d'autres analogies similaires, identiques et parfois différentes des premières de sorte qu'elles enrichissent l'analyse et projettent la recherche vers un domaine autre que le linguistique.

Enfin, l'ouverture du corpus à ces variantes découvertes au sein d'une réalité socioculturelle, historique et même psychologique particulière implique nécessairement un classement méthodique des exemples capable de poser de nouvelles hypothèses pour ce type d'approches. Le corpus est

donc, dans cette perspective, un élément dynamique qui contribue à l'élaboration de la théorie puisqu'il se construit en même temps qu'elle par le biais de la découverte de nouvelles pistes d'investigation servant à faire avancer l'analyse.

Mots-clés : Analogie, classement, contrastif, pragmatique, corpus, discours, empirique, variante/variable.

Références :

CHARAUDEAU, Ph, « Dis-moi quel est ton corpus, je te dirai quelle est ta problématique » *Corpus* [En ligne] 8 | Novembre 2009, mis en ligne le 01 juillet 2010. URL : <http://corpus.revues.org/index1674.html> (consulté le 21 février 2012)

DALBERA, J-P, « Le corpus entre données, analyse et théorie » *Corpus* [En ligne] 1 | Novembre 2002, mis en ligne, le 15 décembre 2003. URL : <http://corpus.revues.org/index10.html> (consulté le 14 février 2012)

GRIZE, J-B, *Logique et langage*, Ophrys 1997.

LACOUR, P, « Discours, texte et corpus », disponible sur : http://alufc.univ-fcomte.fr/pdfs/913pdf_11.pdf

LAKOFF, G, JOHNSON, M, *Les Métaphores Dans La Vie Quotidienne*, Minuit, 1985.

MAYAFFRE, D, « Les corpus réflexifs : entre architextualité et hypertextualité », *Corpus* [En ligne] 1 | Novembre 2002, mis en ligne le 15 décembre 2003, URL : <http://corpus.revues.org/index11.html> (consulté le 12 février 2012.)

MAYAFFRE, D, « Rôle et place des corpus en linguistique : réflexion introductive. *Texto !* [En ligne], Décembre 2005, vol x, n°4. Disponible sur : http://www.revue-texto.net/reperes/Themes/Mayaffre_corpus.html (consulté le 19 février 2012)

MAYAFFRE, D, « Les corpus politiques : objet, méthode et contenu. Introduction. », *Corpus* [En ligne] 4 | Décembre 2005, mis en ligne le 05 septembre 2006. URL : <http://corpus.revues.org/index292.html> (consulté le 12 février 2012.)

MELLET, S, « Corpus et recherches linguistiques », *Corpus* [En ligne] 1 | Novembre 2002, mis en ligne le 15 décembre 2003. URL : <http://corpus.revues.org/index7.html> (consulté le 14 février 2012)

RASTIER, F, « Enjeux épistémologiques de la linguistique de corpus » *Texto !* [En ligne] juin 2004. Rubrique Dits et Inédits. Disponible sur : http://www.revue-texto.net/Inedits/Rastier/Rastier_Enjeux.html (consulté le 13 février 2012)

RICOEUR, P, *La Métaphore vive*, seuil, 1975.

SHEER, T, « Le corpus heuristique : un outil qui montre mais ne démontre pas », *Corpus* [En ligne], 3 | décembre 2004, mis en ligne le 02 décembre 2005. URL : <http://corpus.revues.org/index210.html> (Consulté le 23 février 2012.)

Quel corpus pour étudier les néologismes de la langue de l'économie en Algérie ?

Soufiane LANSEUR

Linguistique, Université de Béjaïa, Béjaïa, Algérie.

E-mail : slanseur@yahoo.fr

Les études linguistiques, en général, et les études lexico-sémantiques qui sont menées sur le français en Algérie s'appuient presque toutes sur des corpus. Le français en tant que première langue étrangère utilisée dans cette aire géographique bénéficie d'une large diffusion dans les différents supports médiatiques : une dizaine de journaux, une chaîne radiophonique, une chaîne de télévision. L'actualité économique prend une large part de l'information diffusée par ces médias. Nous avons eu l'occasion d'effectuer une recherche sur un corpus constitué à la fois d'émissions radiophoniques au nombre de douze et qui durent chacune 45 mn. Ce qui représente 9 heures de discussion en direct. A ce corpus radiophonique oral, parce qu'il s'agit de débats sur des questions économiques avec la participation des auditeurs de la chaîne radiophonique, nous avons rajouté 200 articles de presse relevés dans un hebdomadaire spécialisé dans l'économie. La collecte de ce corpus et son analyse ont été menées dans le but d'étudier l'évolution du vocabulaire de l'économie en Algérie, étant donné que l'Algérie était économiquement socialiste après l'indépendance, mais que la politique économique a été réformée au cours des années 1990 à cause de la crise qu'il l'a secouée durant des décennies.

1. PROBLEMATIQUE

Après le constat de l'existence de divergences entre les définitions données dans les dictionnaires de l'économie et les définitions qu'on trouve dans la presse à propos de concepts et de notions de l'économie, nous avons pris la décision de mener une recherche pour déterminer la néologie dans le discours de l'économie en Algérie. Nous nous sommes donc demandé si le discours utilisé par les médias peut constituer un corpus pour étudier, décrire et analyser les néologismes de la langue de l'économie. Peut-on confectionner un dictionnaire spécialisé à partir de ce corpus ? Cette étude menée à la fois sur le plan formel pour détecter les nouveaux mots qui intègrent le vocabulaire économique, et sur le plan sémantique pour déterminer les nouveaux sens assignés aux mots existant déjà dans les dictionnaires.

2. RESULTATS

a) En effet, le corpus utilisé révèle un décalage entre les définitions données dans les dictionnaires et l'usage de ces mêmes mots dans le discours de l'économie à travers la presse et la chaîne radiophonique. Cette divergence vient du fait que les dictionnaires sont rédigés par des lexicographes et des spécialistes qui ne prennent pas toujours en charge les différences d'usage du français en France et en Algérie.

b) Ce corpus est suffisant pour rendre compte de la plupart des phénomènes néologiques qui se trouvent dans le discours de l'économie parce que les médias ont recours à des spécialistes de l'économie quand ils rédigent leurs articles ou ils réalisent leurs émissions radiophoniques.

c) Dans le contexte algérien, il y a la nécessité de confectionner un dictionnaire de l'économie qui part de la réalité socio-économique de cette aire géographique. L'application des principes de l'économie du marché ne se fait pas de la même manière que dans des pays capitalistes à cause de son passé socialiste. Donc, ce corpus qui révèle un réemploi de concepts importés peut être intéressant pour revoir les notions en relation avec le passé de chaque région géographique.

REFERENCES BIBLIOGRAPHIQUES

- CHANSOU M. (1997), « Méthodologie de la constitution du corpus », Terminologies nouvelles « Enquêtes terminologiques » n° 16, juin 1997.
- FAULSTICH E. (1996), « Spécificités linguistiques de la lexicologie et de la terminologie. Nature épistémologique », Meta, le journal des traducteurs, vol. 41, n° 2, p. 237 -246.
- MEUNIER J -G. (1990), « Le traitement et l'analyse informatique des textes », ICO - Québec, vol 2, n°3.
- RASTIER F. (2005), « Enjeux épistémologique de la linguistique de corpus », La linguistique de corpus , dir. G. WILLIAMS, Presses universitaires de Rennes, Rennes, p 31 -45.

Un corpus numérique d'arabe contemporain : bilan et perspectives

Catherine Pinon*

* Linguistique arabe, IREMAM, Aix-en-Provence, catha@free.fr

Dans le cadre de ma thèse qui visait à étudier le fonctionnement d'un verbe en arabe contemporain, j'ai élaboré un corpus numérique d'arabe écrit contemporain. La constitution du corpus, la collecte des données et leur traitement a nécessité une formation méthodologique et technique particulière dont je tirerai ici les principales conclusions.

Dans cette communication, je souhaiterais d'abord revenir sur les choix qui ont présidé à la sélection des données. J'ai choisi de collecter des données provenant de 7 pays différents et ressortissant à trois genres (la presse, la littérature et les blogs). Ces choix ont été contraignants mais formateurs car ils m'ont amenée à raisonner mon travail : quelle taille devait atteindre mon corpus pour être à la fois représentatif et exploitable ? Sur quels critères sélectionner les données ? Comment extraire automatiquement toutes les occurrences qui m'intéressaient ?

Travaillant sur la langue arabe, des difficultés techniques supplémentaires sont survenues, qui peuvent en partie expliquer qu'aucun corpus de référence n'existe pour cette langue. J'ai travaillé avec le logiciel Lexico 3 pour extraire plus de 15000 occurrences de ce verbe dans mon corpus d'1,5 million de mots. Je souhaiterais présenter les principales difficultés du traitement automatique de l'arabe, dues à ses caractéristiques morphologiques mais aussi à la qualité des textes sources, pour proposer quelques solutions pour améliorer le traitement automatique de corpus en arabe.

Enfin, je souhaiterais présenter certains résultats de mon travail, pour montrer quelle place le contexte a tenu dans l'interprétation des énoncés et comment le lien a pu être fait entre l'objet-corpus, constitué dans un but bien précis, et la langue qu'il représente. Il sera aussi intéressant de montrer comment l'exploration de ce corpus a permis de faire émerger des problématiques totalement nouvelles.

Il s'agira donc, dans cette communication, de dresser le bilan d'une expérience de corpus aboutie, allant de la genèse du projet (le besoin d'un corpus), à sa constitution et à son exploitation, pour une langue dont le traitement automatique est loin d'être évident.

REFERENCES BIBLIOGRAPHIQUES

- Abdelali A., Cowie, J. et Soliman, H. S. (2005), « Building a Modern Standard Arabic Corpus », Workshop on Computational Modeling of Lexical Acquisition .
- Al-Sulaiti L., Atwell, E. (2006), « The Design of a Corpus of Contemporary Arabic (CCA) », International Journal of Corpus Linguistics , vol. 11, p. 135 -171.
- Habash N. (2010), Introduction to Arabic Natural Language Processing , Morgan & Claypoll, 167 p.
- Habert B., Fabre C., Isaac, F. (1998), De l'écrit au numérique. Constituer, normaliser et exploiter les corpus électroniques , Masson, Paris, 342 p.
- Hamdani A., Lachhab K., Erradi, M. (2007), Traitement automatique de la langue arabe , Institut d'Etudes et de Recherches pour l'arabisation , 335 p. + 103 p .
- Lebart L., Salem A. (1994), Statistique Textuelle , Dunod, Paris, 342 p.

