

Exploration textométrique dans le paysage plurilingue de la presse algérienne francophone

Zina Icheboudene ¹, Margareta Kastberg Sjöblom ²

2LASELDI EA2281, Université de Franche-Comté, UFR SLHS, 30, rue Mégevand, 25030 Besançon cedex, zina.icheboudene@univ-fcomte.fr

2ATST, Centre Jacques Petit, EA3187, Université de Franche-Comté, UFR SLHS, 30, rue Mégevand, 25030 Besançon cedex, margareta.kastberg@univ-fcomte.fr

Abstract

The purpose of this paper is to perform a study with focus on the Algerian Francophone press in a sociolinguistic perspective.

The sociolinguistic situation in Algeria is characterized by the coexistence and the constant exchange between several languages: Standard or Literary Arabic, Regional Arabic (Algerian dialects), Berber Languages and French. The presence of two or more languages within one single communication occasion or in the same article is a characteristic of today's Francophone Algerian press.

The study is carried out exploring a large contemporary corpus extracted from three Algerian daily newspapers edited in French during six months. The extraction of loanwords and the identification of all lexical items issued from the different local languages in the corpus is an important step in this study and are carried out with the new tool *Ictena*. The exploration with the *Hyperbase* tool allows the extraction of lexical forms and tokens, statistically specific for this corpus, as well as isotopy correlations, networks and constellations, and makes it possible to observe, at a semantic and at a lexical level, the various associations between the different items of the corpus. This exploration makes it possible, not only to observe and to distinguish the thematic fields of the different plurilingual components of the corpus, but, furthermore, to characterize and describe the characteristics of the code-switching procedure.

Résumé

Cette étude s'intéresse à l'étude de la presse francophone en Algérie dans une perspective sociolinguistique. La situation sociolinguistique en Algérie est marquée par la coexistence et le contact sur le terrain de plusieurs langues : l'arabe littéraire, l'arabe dialectal (algérien), le berbère avec ces différentes variétés et le français. La présence de deux ou même de plusieurs langues à l'intérieur d'un même discours est une caractéristique de la presse algérienne francophone, où le contact entre les langues est omniprésent.

L'objectif est l'analyse d'un corpus de presse algérienne francophone et l'étude de l'impact du plurilinguisme. Il s'agit plus précisément de la constitution d'un grand corpus de presse algérienne francophone contemporaine, de la détection et du recensement de tous les items lexicaux provenant des idiomes locaux dans le corpus.

L'analyse textométrique, l'exploration textuelle et l'analyse statistique permettent d'étudier la distribution de ces emprunts, leurs corrélats et leurs contextes. L'extraction automatique des collocations et des univers cooccurrence permet aussi d'étudier les constellations thématiques de ces emprunts dans les diverses situations d'interférence linguistique. Les résultats permettent de caractériser de manière précise une typologie de discours bien particulier dans un contexte plurilingue.

Mots-clés : analyse de discours, corpus, presse, distribution lexicale, emprunts, contact des langues.

1. Introduction

La situation sociolinguistique en Algérie est marquée par la coexistence et le contact sur le terrain de plusieurs langues : l'arabe littéraire, l'arabe dialectal (algérien), le berbère avec ses différentes variétés et le français. En effet, la langue française est toujours présente dans la société algérienne, aussi bien à l'oral qu'à l'écrit. Elle est même perçue comme la langue du savoir et de l'ouverture. Le français est toujours la langue véhiculaire dans certains secteurs-clés de la société en l'occurrence : l'administration, l'économie, l'enseignement supérieur et certains médias.

Le contact de la langue française avec les langues locales en Algérie a donné lieu à des situations d'interférences linguistiques qui se traduisent par plusieurs phénomènes : les emprunts lexicaux, l'alternance codique, les nouvelles formulations syntaxiques, etc.

La présence de deux ou même de plusieurs langues à l'intérieur d'un même discours est notamment une caractéristique de la presse algérienne francophone, où le contact entre les langues est omniprésent.

L'évolution de la presse écrite en Algérie est liée à l'évolution du système politique du pays : à partir de 1962¹ le pouvoir en place a exclu toute presse indépendante et libre ou d'opinion partisane. L'adoption d'une constitution pluraliste en février 1989 (Cheurfi, 2008), et la levée du monopole gouvernemental sur la presse écrite ont provoqué l'émergence d'une presse privée et partisane. Cette dernière considère que le nationalisme et le patriotisme ne constituent pas un frein à l'ouverture de l'Algérie sur le monde et à une reconnaissance du plurilinguisme.

À partir des années 90², avec le changement du code de l'information, on constate une mutation profonde de la presse écrite, avec la création de dizaines de quotidiens, d'hebdomadaires et de mensuels de statut privé édités en arabe et en français. Selon les statistiques officielles de 2008 33 titres sont publiés en français et 32 en arabe.

L'objectif de notre étude est l'analyse d'un corpus de presse algérienne francophone et l'étude de l'impact du plurilinguisme. Il s'agit plus précisément de constituer un grand corpus de presse algérienne francophone contemporaine, de détecter tous les items lexicaux provenant des idiomes locaux (l'arabe et ses variétés, ou le berbère et ses variétés) dans les discours de cette presse et de les recenser avec l'aide de plusieurs logiciels pour pouvoir les analyser dans leurs contextes.

Le logiciel textométrique, avec ces fonctions statistiques et hypertextuelles, permet de situer les apparitions de ces emprunts dans le corpus et de comparer les différentes variantes morphosyntaxiques. L'extraction automatique des univers cooccurrentiels et des collocations permet aussi d'étudier les constellations thématiques de ces emprunts dans les diverses situations de leur intégration.

1 La proclamation de l'indépendance de l'Algérie a eu lieu le 3 juillet 1962.

2 Le changement du code de l'information et le lancement d'une presse privée et indépendante datent du 3 avril 1990.

2. Constitution d'un corpus de presse

Le corpus de cette étude est composé de l'ensemble des articles des trois quotidiens algériens généralistes en langue française : *Liberté*, *El-Watan*, et *Le Quotidien d'Oran*. Il comprend les éditions quotidiennes sur une période de six mois, du 08 septembre 2008 au 08 mars 2009, représentant au total 423 éditions.

Le recueil de ces données s'est fait à travers les sites internet de ces quotidiens qui proposent l'édition du jour en format « PDF ». Afin de constituer un corpus exploitable, nous avons procédé en différentes phases de travail : découpage du journal en articles à l'aide du logiciel *Adobe Photoshop*, océrisation avec le logiciel *Abbyy FineReader*, pour convertir les documents en deux formats exploitables « txt » et « html ».

Nous avons retenu tous les articles de l'ensemble des rubriques, et avons seulement exclu les pages intégralement réservées à la publicité.

Les textes du corpus sont repartis en trois sous-corpus de tailles relativement homogènes. A l'intérieur de chacun de ces trois corpus se déclinent 141 éditions (correspondant au nombre de parutions sur la période de six mois choisie) et un grand nombre d'articles assez variable d'un journal à l'autre. Les différences dans le nombre d'articles et le nombre d'occurrences s'expliquent par la différence de nombre de pages des éditions, qui varie d'un numéro à l'autre : entre 18 et 30 pages pour *El-Watan*, de 30 à 32 pour le *Quotidien d'Oran*, et de 20 à 26 pour *Liberté*.

Ce corpus a été traité et exploré avec plusieurs logiciels. La segmentation du corpus a été réalisée avec deux logiciels : *ASTADIAG* (pour *Astartex* et *Diatag*)³ et *Ictena*⁴. Une lecture de 423 éditions de ces trois quotidiens à l'aide d'*Ictena* et à travers la totalité de leurs articles, nous a permis de recenser 13.130.531 occurrences.

Journal	Nombre d'éditions	Nombre d'articles	Nombre d'occurrences
<i>Liberté</i>	141	8506	3.569.219
<i>El-Watan</i>	141	12221	4.978.985
<i>Le-Quotidien d'Oran</i>	141	10398	4.582.327
Total	423	31125	13.130.531

Tableau 1 : Le corpus

L'exploration statistique des données a été effectuée avec le logiciel *Hyperbase*, dans sa version récente 9.0. Ces logiciels de traitement quantitatif des grands corpus textuels autorisent un ensemble de traitements sur des corpus de textes prédéfinis et numérisés. L'exploration

3 Environnement développé par J.-M. Viprey, laboratoires **ATST** – Archives, Textes, Sciences des Textes (EA 3183) et **LASELDI** – Laboratoire de Sémio-linguistique, Didactique, Informatique (EA 2281), Université de Franche-Comté, (http://laseldi.univ-fcomte.fr/document/viprey/page_JMV.htm).

4 Interface de Consultation de Textes numériques en vue de l'Analyse, développé par N. Salzard, et Ph. Schepens, **LASELDI** – Laboratoire de Sémio-linguistique, Didactique, Informatique (EA 2281), Université de Franche-Comté, <http://laseldi.univ-fcomte.fr/ictena/DOC/documentation.php>.

statistique donne la possibilité d'analyses statistiques diverses, non seulement traditionnelles, mais également plus spécifiques et qui nous semblent convenir à notre étude précise, ce qui motive l'utilisation complémentaire de plusieurs logiciels.

L'analyse des spécificités lexicales aussi bien d'un point de vue endogène qu'exogène (en comparaison avec données du corpus *Frantext*), le calcul des cooccurrences et l'extraction des corrélats thématiques constituent en effet des outils précieux dans une étude de caractère sociolinguistique. Etant donné l'aspect d'interférence entre items lexicaux français et d'autres en arabe et en berbère, nous avons choisi de travailler sur les formes graphiques et de ne pas avoir recours à la lemmatisation.

3. Détection des items lexicaux étrangers à la langue française

Afin de détecter les items lexicaux étrangers à la langue française dans notre corpus, nous l'avons confronté à un dictionnaire de référence, celui de *DiaTag*, strictement dédié au français moderne et contemporain (Lethier et Viprey, 2008). *Diatag* s'intègre dans l'environnement *Astartex*, dédié à l'exploration assistée elle-même, et aux calculs statistiques.

L'application et l'analyse de *DiaTag* nous ont permis l'identification initiale des items inconnus dans le corpus. En se référant au dictionnaire nous avons identifié 133.633 formes et 536.801 occurrences inconnues, c'est-à-dire environ 4.1% du volume total. Ces formes correspondent à des mots que le dictionnaire du logiciel *Diatag* ne connaît pas ; il s'agit de :

- mots étrangers (arabe, berbère, anglais, espagnol,...),
- néologismes,
- noms propres.

Naturellement, on trouve aussi des fautes et des coquilles, ce problème provient de l'architecture même des articles de presse en colonnes, de tailles et de positions différentes, en format PDF, qui nécessite le découpage et à l'océrisation des articles.

Afin d'obtenir la liste des items lexicaux en langues locales d'Algérie présents dans notre corpus, nous avons procédé au nettoyage de la liste et avons sélectionné tous les mots issus de l'arabe et du berbère : les emprunts, les néologismes et les termes hybrides.

En effet, nous avons relevé plusieurs types de mots provenant des langues locales. Principalement, il s'agit de tous les emprunts de l'arabe ou du berbère, comme *chouhada* (martyrs) *moudjahiddine* (combattants), *hogra* (injustice, abus) et *Kafala*⁵, etc. En outre, nous avons relevé des items hybrides qui portent des marques de dérivations morphologiques, syntaxiques ou sémantiques, des mots en langue locale « contaminés » par le français (préfixation, suffixation, adjonction de déterminants, de marque de genre et de nombre, etc.), par exemple : *taxieur* (pour chauffeur de taxi), *ramadhanesques* (relatif au *ramadhan*, exemple : les soirées *ramadhanesques*), *ex-wali* (ex-préfet), etc.

Nous avons pu remarquer que ces items détectés dans le corpus font souvent référence à des champs sémantiques bien caractéristiques, renvoyant à l'identité algérienne.

5 Loi musulmane qui précise et régleme les conditions légales pour l'adoption d'enfants abandonnés.

4. Classification des mots empruntés ou dérivés

A la suite de la segmentation de *Diatag* nous avons constitué un inventaire des mots empruntés qui font référence à différents champs sémantiques bien spécifiques ; les termes les plus représentatifs font référence à :

- la religion musulmane : *zakate*⁶, *fetwas*⁷, *tarawih*⁸, *kafala*, *rahma*⁹, *zaouïas*¹⁰, *iftar*, etc., sont des emprunts qui désignent les pratiques religieuses.
- les administrations et institutions : *daïra*, *wali*, *wilaya*, *mouhafadha*, *mouhafedh*, *houkouma echaâb*, etc., se réfèrent à la division administrative et aux fonctionnements institutionnels de l'Algérie.
- l'espace et les lieux : *dechra*, *houma*, *hay*, *douar*, *madina*,
- l'art, la culture et les traditions : *chaâbie*, *hawzi*, *aïssaouia*, *melhoun*, *qacidate*, *naili*, *zorna*, *maqam*, *qaâda*, *seroual*, *zerda*, *meïda*, *kheïma* etc., regroupent tous les emprunts relatifs aux chants traditionnels, aux styles musicaux, aux tenues vestimentaires et aux différentes pratiques culturelles
- la gastronomie : *zlabia*, *frik*, *zitoune*, *h'rira*, *diouls*, *baklaoua*, *bourek*
- la réalité sociale : *harraga*, *harga*, *hogra*, *khouya*, *zawali*, *guellil* etc., viennent pour la plupart du langage populaire (dialectes arabe et berbère).

En examinant cet inventaire, nous avons pu constater que les journalistes de la presse algérienne francophone ont souvent recours aux langues locales pour désigner ou dénoter une réalité propre à l'environnement social et culturel qui caractérise la société algérienne.

La distribution de ces emprunts n'est pas régulière dans le corpus. La distribution des fréquences de ces termes à l'intérieur de notre corpus constitué des trois quotidiens, calculée par le logiciel *Hyperbase*, permet de constater des différences notables. Durant le premier mois de notre corpus (septembre 2008), nous notons en effet une fréquence importante des emprunts relatifs à la religion musulmane, à la gastronomie, à l'art et à la culture algériens. Il s'agit en effet de la période de *ramadhan*¹¹, qui s'étale sur trente jours.

La fréquence de ces termes empruntés est également à noter pendant deux autres mois de la période observée, bien qu'elle reste en deçà du seuil significatif de 5%. Il s'agit premièrement du mois d'octobre 2008, le mois où les Musulmans fêtent la journée de l'*Aïd*¹² et deuxièmement du mois de mars 2009 quand a eu lieu la fête religieuse appelée *Mawlid an-Nabawi*, jour d'anniversaire de la naissance du Prophète Mohamed¹³.

6 Aumône distribuée à l'occasion de la fin du ramadhan aux pauvres.

7 Discours d'un spécialiste de loi musulmane portant un avis religieux.

8 Prière facultative exécutée chaque soir du Ramadan.

9 Charité, pitié, clémence.

10 Édifice religieux musulman.

11 Mois lunaire pendant lequel les musulmans doivent s'astreindre à l'abstinence (nourriture, boisson, tabac, relations sexuelles) entre le lever et le coucher du soleil.

12 Fête religieuse musulmane consacrant la fin du jeûne du mois de *ramadhan*.

13 Douzième jours du troisième mois du calendrier musulman.

Ces périodes sont marquées par des pratiques religieuses particulières et par un changement dans les habitudes culinaires, sociales et économiques des Algériens, d'où l'importance accordée par les journalistes à ce vocabulaire emprunté aux langues locales à l'intérieur d'un discours en langue française.

Une grande partie des items arabes font en effet référence à la religion musulmane. La distribution de fréquences relatives de l'ensemble de termes : « *Tarawih* », « *Zakat* »¹⁴, « *Kafala* », « *fatwa* », « *rahma* » reflète très bien l'affluence de ces termes dans la partie initiale de la période étudiée :

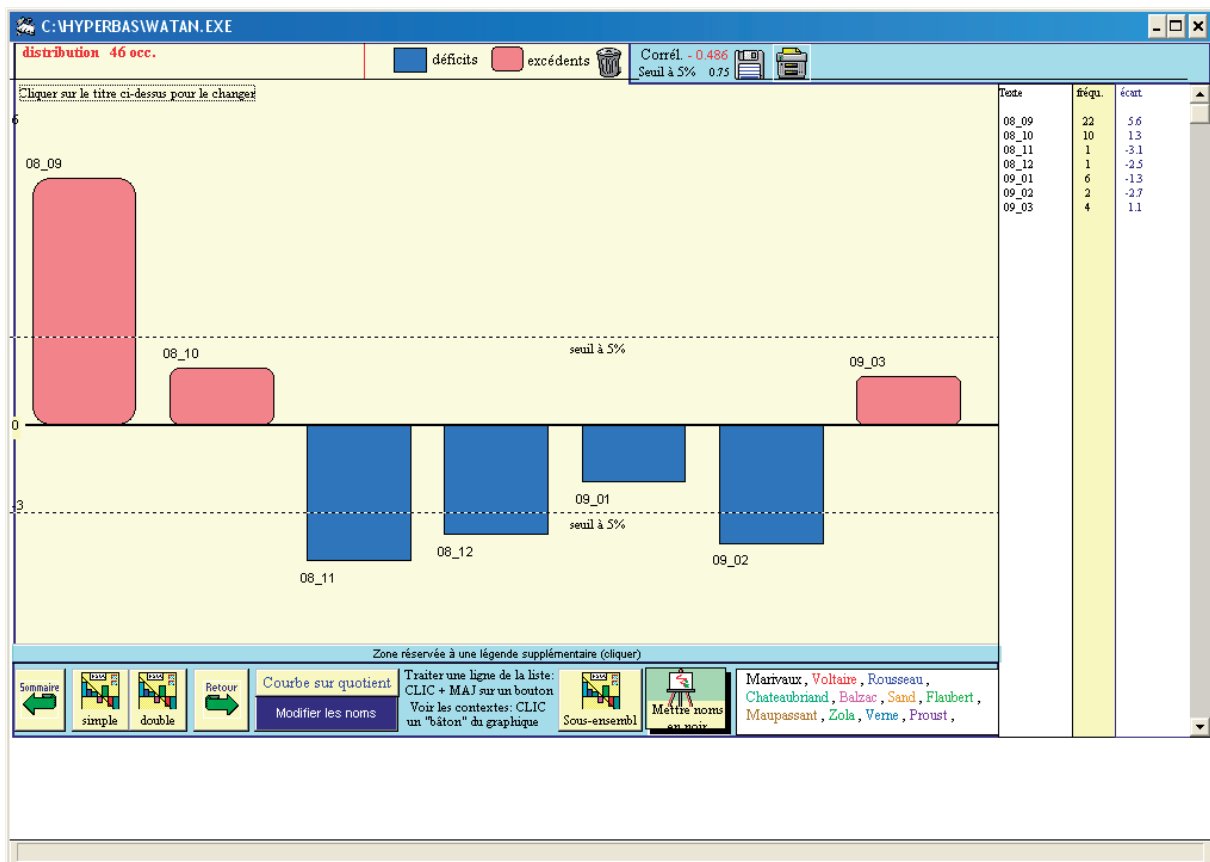


Figure 1 : Distribution des fréquences relatives des termes religieux dans El-Watan

La motivation de l'introduction de ces emprunts dans le discours français est assez claire : les journalistes ont besoin d'intégrer des mots pour exprimer des notions non répertoriées dans la langue française, ou bien pour désigner des références propres à la culture arabo-musulmane. Toutefois, non seulement l'apparition d'emprunts isolés dans le corpus est caractéristique de ce corpus, mais l'environnement lexical qui les entoure l'est également.

14 Aumône distribuée à l'occasion de la fin du ramadhan aux pauvres.

5. Cooccurrences

L'apparition de lexies empruntées est souvent très liée à des contextes particuliers ; prenons par exemple un emprunt à l'arabe littéraire; le mot *rahma*. Il s'agit d'une notion religieuse qui désigne la charité, la clémence et la compassion. L'analyse des contextes autour des apparitions de ce terme nous permet de constater qu'il est très fréquent pendant la période du *ramadhan*.

Dans la société algérienne *rahma* est aussi utilisé pour qualifier l'esprit généreux des pratiquants qui s'investissent dans la création des restaurants appelés « *matham el rahma* », correspondant aux « restaurants du cœur » en France. Ces restaurants sont considérés comme un refuge pour les pauvres et les personnes sans familles.

L'analyse de la cooccurrence des deux items montre une forte corrélation des deux occurrences. La figure ci-dessous montre des exemples en contextes.

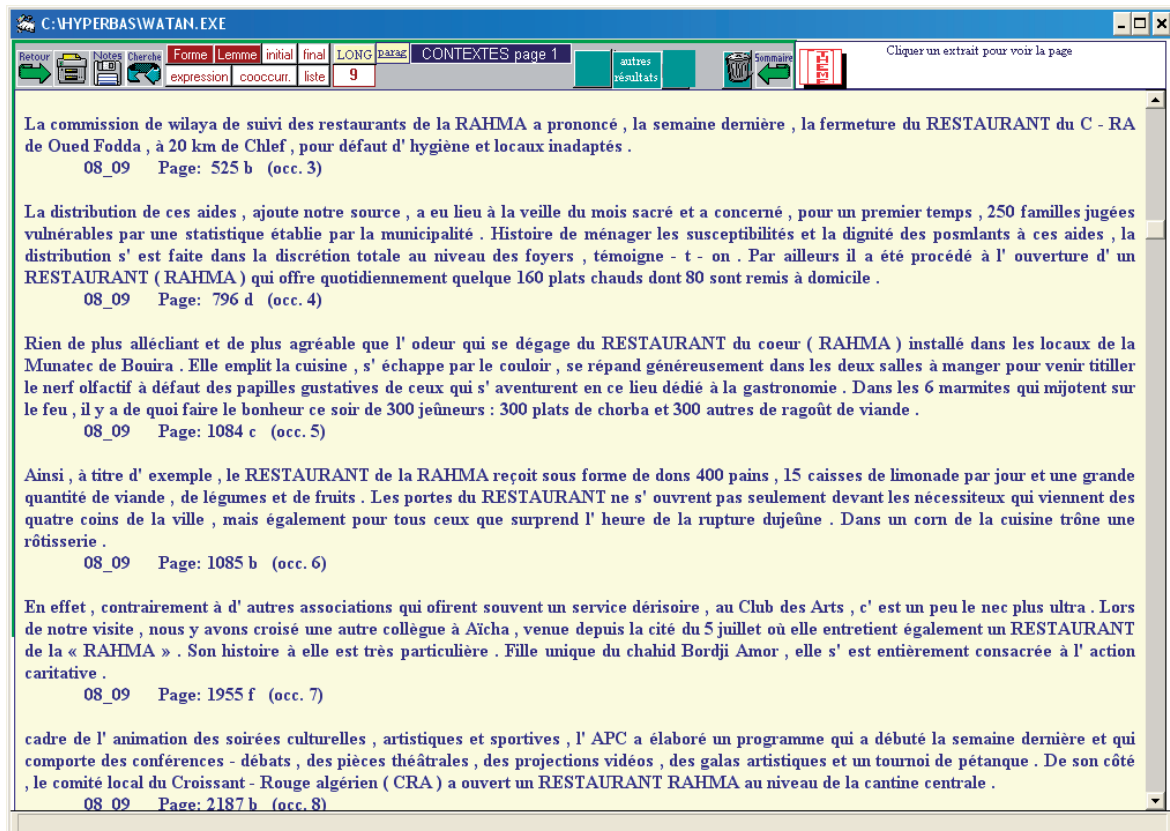


Figure 2 : Contextes de la cooccurrence *rahma* et restaurant dans *Liberté*

Ces exemples illustrent bien la lexicalisation d'une constellation figée en deux langues. Le *restaurant* en français et le *rahma* en arabe littéraire prennent dans le français algérien aujourd'hui un sens particulier.

Ces collocations et lexicalisations des deux items de deux langues différentes sont fréquentes et très caractéristiques de la presse algérienne francophone. Le mot *rahma* ici renvoie à une réalité religieuse et sociale liée, dans la plupart de ses apparitions, au mois de *ramadhan*, que l'expression « resto du cœur » ne pourrait refléter ou intégrer.

En outre, dans ce contexte, le phénomène d'emprunt du mot *rahma* se manifeste comme un ancrage pour marquer l'identité culturelle et religieuse du pays à travers les discours de la presse. Il est également intégré dans ces discours pour caractériser le mois de *ramadhan* comme période d'aide entre la population et le nommé « mois de *rahma* » :

Quel que soit le degré de la faute, nous estimons que les sanctions sont très disproportionnées et surtout en plein mois de la rahma où doivent se multiplier les gestes de solidarité. (Liberté, le 08/10/2008)

6. Associations thématiques

L'utilisation d'emprunts arabes ou berbères, omniprésents dans le discours de la presse algérienne francophone, a en effet une forte valeur identitaire. Nous avons également relevé un nombre important d'emprunts dont l'équivalent existe dans la langue française ; tel est le cas des mots *hogra* (injustice ou abus), *chahid* (martyr), *chabab* (jeunes), *khaiïma* (tente ou guitoune) ; *dechra* (village) et bien d'autres.

Pour illustrer ce phénomène, nous avons choisi l'exemple du mot de l'arabe dialectal « *hogra* ». Cet item nous semble intéressant, non pas par sa fréquence relativement limitée (88 occurrences), mais par son emploi bien particulier dans le discours journalistique algérien et des « moments discursifs » (Moirand, 2007) bien particulier. En effet, l'analyse de l'environnement lexical qui entoure l'item et la comparaison de cet univers lexical avec celui d'une de ses équivalents dans la langue française « *injustice* » permet de constater des différences notables.

Au-delà de l'étude de la simple cooccurrence, le logiciel Hyperbase permet l'extraction d'une association thématique, d'un microcosme cooccurentiel d'une autre manière. Grâce à l'extraction automatique du contexte d'un item lexical, ici *hogra*, le programme crée un sous-corpus qui est soumis à un calcul de spécificité, puisqu'on ne recherche plus une relation entre un mot et un texte, mais une relation privilégiée entre les mots eux-mêmes - ce que mesure aussi le calcul de corrélation. Cependant la procédure ne se réduit pas ici à deux mots confrontés, mais à l'ensemble indéfini de tous les mots qui peuvent se trouver dans l'entourage d'un mot (ou d'un groupe de mots) qu'on définit comme étant le pôle, comme ici *hogra*.

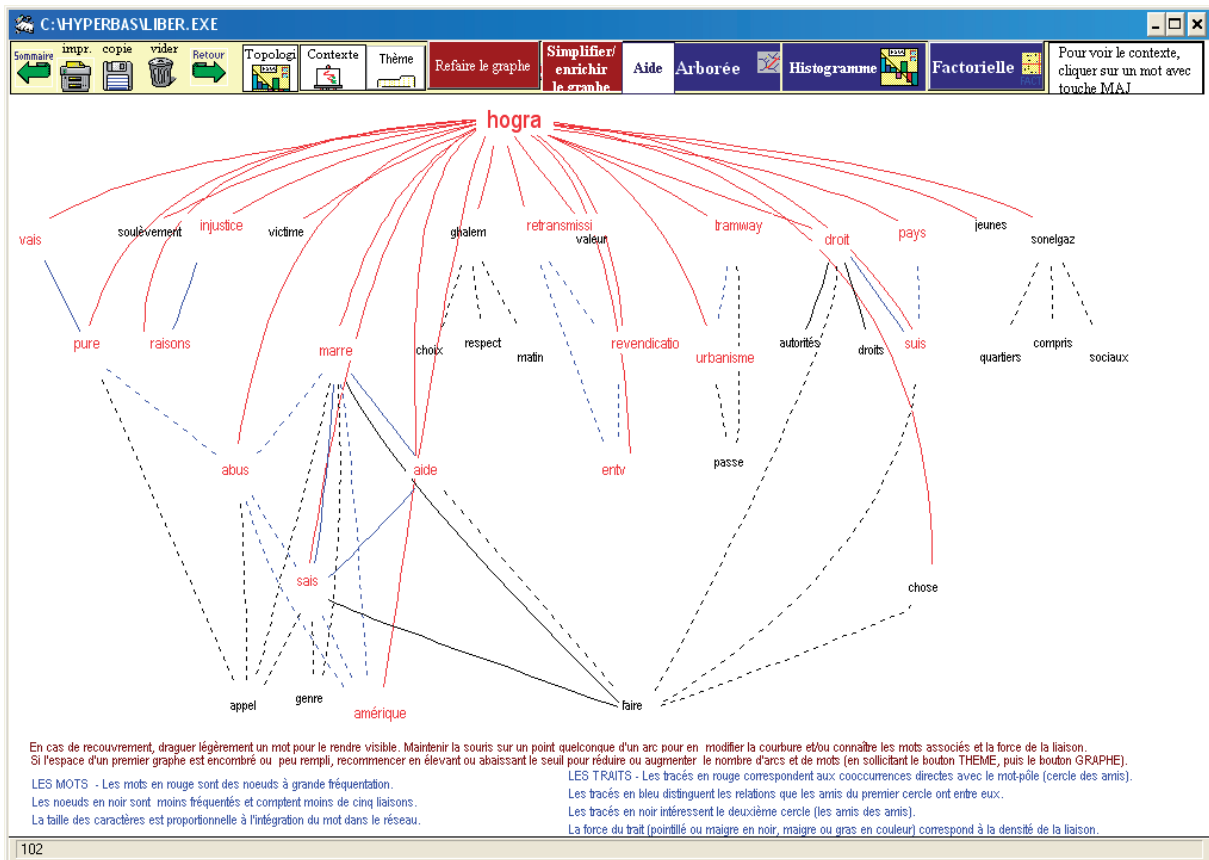


Figure 3 Association thématique à partir du pôle hogra

Les mots en rouge correspondent ici à des nœuds de forte fréquentation, et ceux en noir à des nœuds moins fréquentés, n'ayant pas de contact direct avec le mot-pôle. Les traits gras correspondent aux cooccurrences directes avec le pôle, et les traits fins aux cooccurrences indirectes, c'est-à-dire en cooccurrence avec les cooccurrents.

Les liens les plus forts du *hogra* sont *pays*, *tramway*, *jeunes*, *marre*, *injustice*, *chômage*, *soulèvement* et *revendication*. Le graphique permet aussi de recenser les liens qu'entretiennent les cooccurrents avec d'autres items lexicaux, *hogra* étant lié à l'*abus*, à l'*aide* etc.

La même analyse de l'équivalent français *injustice* montre des résultats assez différents. Ici le calcul est effectué de la même manière que dans le calcul de tout à l'heure, c'est-à-dire le paragraphe entourant le pôle *injustice*.

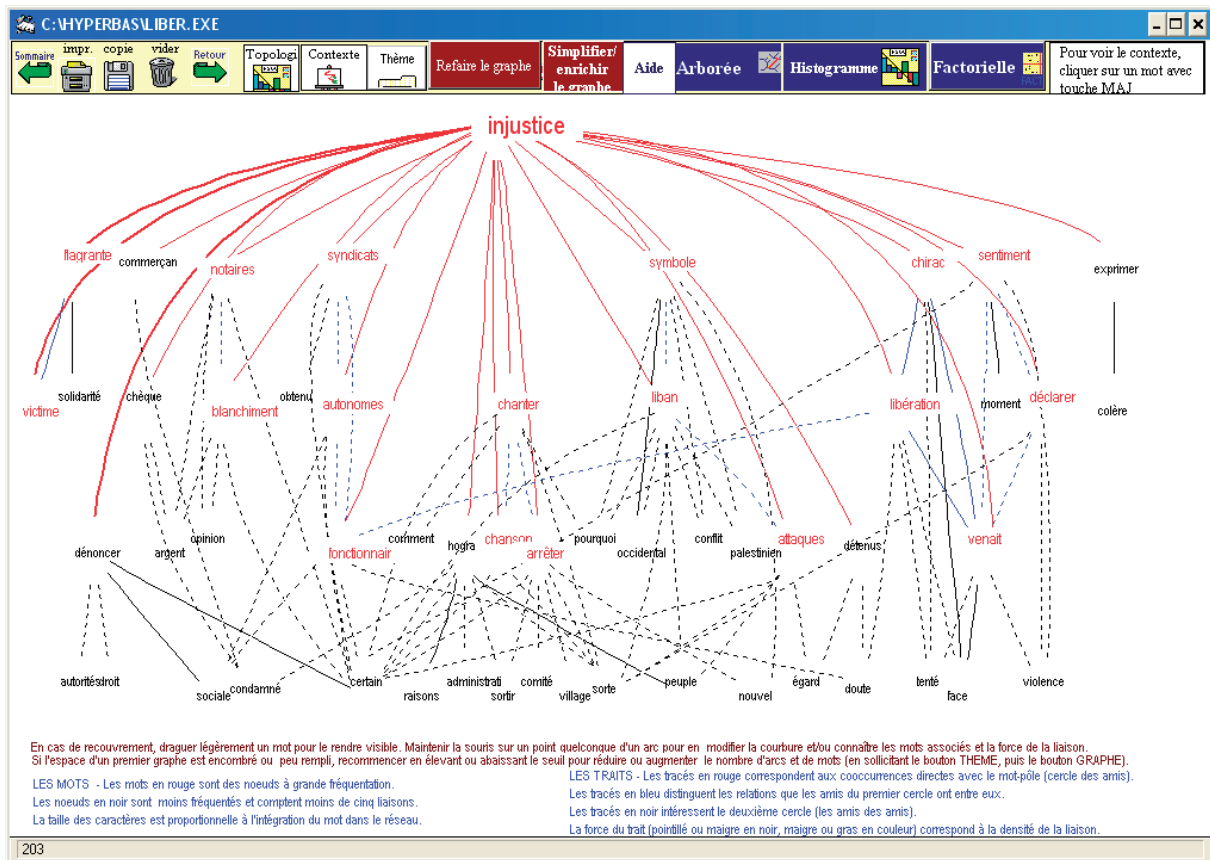


Figure 4 : Association thématique à partir du pôle injustice

Les cooccurrents les plus forts de *injustice* sont *flagrante*, *syndicats*, *Chirac*, *notaires*, *commerçant* et *Liban*. Autour du mot *Liban* gravitent les termes faisant référence à l'actualité politique internationale : *Palestinien*, *occidental*, *conflit* et *attaque*. C'est à peu près la même constellation cooccurrentielle que l'on obtient lorsque l'on analyse la presse française contemporaine.

En revanche la constellation thématique qui gravite autour de « *horga* » est fortement connotée. Le terme est généralement employé dans les articles traitant les problèmes de caractère social en Algérie.

Il est fréquemment employé pour désigner l'arbitraire des décisions officielles et les abus d'autorité. Le terme dénote en effet le mépris des pouvoirs publics envers les jeunes :

« Selon certains témoignages, la gestion douteuse des postes entrant dans le cadre du social, la marginalisation et la **horga** étaient à l'origine de cette vague de violence ». (*El Watan*, le 05/10/2008)

« Elles avaient pour toile de fond le chômage, la **horga**, des jeunes revendiquaient leur droit ». (*Liberté*, le 23/02/2009)

Les journalistes ont aussi recours à ce terme pour définir l'injustice envers les femmes :

« Le contrat de mariage pour éviter la **horga** » (*El Watan*, le 28/10/2008)

« *Les associations féminines s'échinent à dénoncer la **hogra**, les violences et les exactions quotidiennes...* » (*Le Quotidien d'Oran*, le 08/03/2009)

L'émigration clandestine est un autre problème d'actualité en Algérie. Le mot *harraga*, qui signifie « émigré clandestin », est un terme populaire qui fait parti du nouveau lexique émergé dans la société algérienne ces dernières décennies.

Il s'agit d'un néologisme qui vient de l'arabe, créé à partir du verbe « brûler » qui donne le nom « les brûleurs ». Il est employé pour qualifier ces jeunes algériens qui « brûlent leurs papiers », « leurs passé », et qui traversent les frontières clandestinement pour rejoindre les côtes européennes. C'est un terme qui a également attiré notre attention par sa haute fréquence dans le corpus : 594 occurrences. Dans le même contexte nous avons relevé le terme *harga*, c'est ainsi qu'on appelle « le fait de fuir » le pays clandestinement. Il est fort intéressant de constater la cooccurrence des termes *hogra* et *harga* qui ensemble acquièrent un sens particulier et une valeur quasi-proverbiale.

Nous avons relevé dans le corpus les exemples suivants où les journalistes jouent sur la sonorité des deux mots « *hogra* » et « *harga* » :

« *Immigration Entre **hogra** et **harga*** » (Immigration entre injustice et fuite)
(*Le Quotidien d'Oran*, le 02/03/2009)

« *Que diraient nos valeureux « chouhada » s'ils pouvaient revenir et voir les nouveaux maquis de la désespérance, de la drogue, du suicide, de la « **hogra** » et de la « **harga** » pour un quignon de pain, du banditisme en col bien blanc ?* » (*Le Quotidien d'Oran*, le 15/10/08)

Les journalistes emploient en effet ces deux termes *harga* et *hogra* pour donner une dimension sociopolitique aux situations complexes auxquelles est confrontée la jeunesse algérienne, « le mépris », « l'injustice » et « l'émigré clandestin ». Ces jeunes se retrouvent donc devant un choix difficile : soit continuer à vivre dans le pays, soit aller à la recherche d'une porte pour fuir.

Le choix des journalistes d'employer des termes propres à la réalité sociale algérienne peut être considéré comme une stratégie pour s'approcher du lecteur algérien (ouvrier- chômeur- retraité- jeunes diplômés), préoccupé par les problèmes du quotidien, et touché par la crise sociale qui traverse le pays.

En ayant recours à des mots comme *hogra* et *harga*, les journalistes participent à la création d'un code spécifique que seul un lecteur algérien peut déchiffrer, et qui échappe aux autres lecteurs francophones.

7. Variations morphosyntaxiques

Une autre caractéristique très particulière de ces items lexicaux est la variabilité de leurs transcriptions graphiques. En effet, aucune forme n'est retenue comme la seule et unique pour la transcription des mots empruntés ou des néologismes dialectaux. Cela peut se traduire par l'absence d'une norme, comme l'indique un des journalistes de quotidien *El Watan* lors d'un entretien « *nous ne possédons pas une charte orthographique qui nous oriente dans la transcription de ces mots qui ne sont pas intégrés dans la langue française* ».

Les emprunts à l'arabe sont généralement transcrits littéralement. Certains emprunts prennent deux ou même plusieurs formes graphiques, c'est le cas de *tarawih*, *tarawihs* et *taraouih*, *fetwa* et *fatwa*, *hay* et *hai*, etc. Cette variabilité des transcriptions peut être expliquée par les caractéristiques d'intégration phonologique qui peuvent varier considérablement en Algérie.

Ces variations constituent une grande difficulté dans l'analyse lexicométrique et mettent parfois en cause la fiabilité des résultats, de la même manière que dans les études statistiques des corpus en ancien français (Brunet, 2009). Cependant, l'étude des variantes graphiques ouvre la voie à des études sociolinguistiques et dialectales et permet de constater que ces variations sont souvent liées à des différences dialectales entre l'Ouest et l'Est, et à des origines arabophones ou berbérophones des locuteurs du pays ainsi qu'à la grande disparité sociale.

On constate ce même manque d'homogénéité dans l'usage du pluriel, qui est indiqué différemment en arabe et en français. Nous trouvons parfois un -s indiquant un pluriel et parfois un pluriel sans indicateur graphique.

« *Que la décision de l'Etat de pénaliser l'acte d'émigrer clandestinement est absurde quand on sait que des jeunes **harraga** mettent volontairement leur vie en péril en se jetant à l'eau.* »
(*Le Quotidien d'Oran*, le 16/09/08)

« *A Montélimar, précisément - la ville du nougat -, ces **harragas** du passé, sans logement ni nourriture, démunis face au froid, se voient réduits à la pire misère.* » (*El Watan*, le 9/10/08)

8. Conclusion

L'objectif de cette étude n'était pas de traiter tous les aspects de la presse francophone algérienne. Il s'agit ici d'une expérience qui s'intéresse à un aspect spécifique : relever et étudier les emprunts des langues locales dans un corpus en langue française. L'exploration de plusieurs logiciels s'est ici révélée d'une grande utilité. L'exploration d'*Ictena*, interface du dictionnaire du logiciel *Diatag*, a rendu possible le recensement efficace des termes étrangers à l'intérieur du corpus. Le logiciel *Hyperbase*, avec ses fonctions hypertextuelles et statistiques, nous a permis d'étudier l'interférence des langues arabe et berbère ainsi que leurs différents dialectes.

Cette étude montre bien que le processus d'intégration des mots de langues locales dans les discours de la presse algérienne francophone est souvent l'expression de l'évolution sociale. Une grande partie de ces mots traduit les diverses mutations connues par les différentes communautés linguistiques algériennes. En outre, le recours aux termes des langues locales dont les équivalents peuvent exister dans la langue française apparaît comme un procédé d'identification linguistique, un ancrage dans la société et dans la culture algériennes. En effet, la plupart de ces termes traduisent les diverses mutations connues par les communautés linguistiques algériennes. Le recours à l'emprunt devient ainsi nécessité, car parfois le contenu sémantique de leurs équivalents dans la langue française ne couvre pas la même réalité qu'en arabe, ou n'a pas la même profondeur. Ces termes témoignent donc de la présence importante d'une mémoire collective et d'une volonté d'afficher une identité linguistique. Selon Charaudeau (1997), le discours de la presse est sans doute le lieu de « construction d'un miroir social ». Ces mots empruntés fonctionnent discursivement en tant que marques lexicales de la situation socioculturelle algérienne. Les journalistes y ont recours pour véhiculer des représentations propres à l'espace géographique et socioculturel algérien.

Références

- Adam J.-M. (2005). *La Linguistique textuelle. Introduction à l'analyse textuelle des discours*. Paris, Armand Colin.
- Brunet E. (2009). *Compte d'auteurs, Tome 1, Etudes statistiques de Rabelais à Gracq*. Paris, Honoré Champion
- Charaudeau P. (1997) *Le discours d'information médiatique : la construction du miroir social*. Paris, Nathan.
- Cheurfi A. (2008). *La presse algérienne : genèse, conflit et défis*. Alger, Casbah éditions.
- Dubois J., et al. (1973). *Dictionnaire de linguistique*. Paris, Larousse.
- Kastberg Sjöblom M. (2008) "Attirance thématique : fréquences ou séquences ? Extraction des isotopies sémantiques d'un corpus textuel", *5èmes Journées de Linguistique de Corpus*, G. Williams (éd.), Université de Bretagne Sud, Lorient, web.univ-ubs.fr/corpus/jlc5/ACTES/ACTES_JLC07_kastberg.pdf
- Kastberg Sjöblom M. (2010). "Costellazioni tematiche in un corpus letterario italiano", *JADT 2010, 10e Journées internationales d'Analyse statistique des Données Textuelles*, S. Bolasco (éd.), Roma, Edizioni Universitarie di Lettere Economia Diritto
- Lethier V. & Viprey J.-M. (2008). «Annotation linguistique de corpus : vers l'exhaustivité par la convivialité». In Heiden S. et Pincemin B. (éds.) *JADT'09, 9èmes Journées internationales d'Analyse statistique des Données Textuelles*, Lyon, Presses Universitaires de Lyon.
- Moirand S. (2007). *Les discours de la presse quotidienne : observer, analyser et comprendre*, Paris, Presse Universitaire de France.
- Viprey J.-M. (2005). « Philologie numérique et herméneutique intégrative », In Adam J.-M. et Heidmann U. (éds.) *Sciences du texte et analyse de discours : enjeux d'une interdisciplinarité*, Genève, Slatkine.
- Viprey J.-M. (2006) « Structure non-séquentielle du texte ». In *Langages n°163 - Unité(s) du texte*. Paris, Arman Colin.