

Usages de la lexicométrie en analyse de discours

François LEIMDORFER* et André SALEM**

Pour le sociologue, l'anthropologue, l'historien, le psychologue et le linguiste, qui travaillent sur des discours, des textes écrits ou des retranscriptions (questionnaires, interviews, histoires de vie, etc.), pouvoir analyser des corpus étendus est d'un grand intérêt. En effet, l'analyse de discours se fait habituellement sur des fragments, dont on tire des hypothèses que l'on teste sur des discours plus larges.

Le projet d'utiliser des techniques automatisées en matière d'analyse de contenu et d'analyse de discours est déjà ancien. La codification des thèmes en analyse de contenu est une pratique courante, et Michel Pêcheux avait créé l'*Analyse automatique du discours* (AAD) en 1969, à partir d'une approche syntaxique (à la suite de Z. Harris). Ces méthodes exigent cependant une préparation importante du corpus.

L'exigence pour l'analyse est de pouvoir travailler sur des textes en « langue naturelle », avec le minimum de transformations par rapport au discours tel qu'il s'est dit ou écrit. Il est ainsi très important, dans des enquêtes à questions ouvertes, de pouvoir traiter les réponses telles qu'elles ont été formulées, et de ne pas devoir en passer par une pré-codification thématique, qui est déjà l'aboutissement d'une analyse et d'un jugement (ACHARD, 1991). Ceci est bien entendu valable pour toutes les données orales (interviews, histoires de vie, conversations), d'autant que lorsque l'on travaille sur des retranscriptions, la « donnée » a déjà subi de nombreuses modifications : le passage de la situation de face à face à la bande magnétique (problèmes du parasitage, de la qualité de l'enregistrement, de la non-sélectivité des voix et des bruits, de la perte des gestes, des mimiques et des attitudes, etc.) et le passage de l'enregistrement de l'oral à la transcription (perte de la durée des silences, de l'accentuation, de l'intonation, etc.).

* Ingénieur CNRS, Laboratoire tiers monde, Afrique, IEDES-CECOD, 58, bd Arago 75013 Paris.

** Enseignant-chercheur à l'université Paris-III, Lexicométrie et textes politiques, ENS de Fontenay-Saint-Cloud, grille d'honneur du Parc, 92211 Saint-Cloud.

LE LOGICIEL LEXICO

Le logiciel Lexico, mis au point par André Salem, est un traitement « lexicométrique », c'est-à-dire statistique, d'un texte (LEBART et SALEM, 1989 ; SALEM, 1987). Il fonctionne sous Macintosh.

Le texte

Le corpus ne doit pas subir de préparation particulière. De plus, le transcritteur peut y faire les transformations qu'il souhaite : indiquer le nom des locuteurs, adopter une norme pour certaines expressions ou mots étrangers, inscrire les pauses, les reprises, les intonations par un signe particulier, indiquer des variables « extra-discursives » (tout type de variable connue de lui), etc.

Le texte ne sera pas *lemmatisé*¹. Les formes apparaîtront dans le traitement lexicographique telles qu'elles ont été saisies : une forme au singulier et au pluriel comptera pour deux formes différentes, de même qu'un adjectif au masculin et au féminin ou qu'un verbe sous ses différentes formes conjuguées.

On distribue le corpus en plusieurs parties, en introduisant (par une simple édition du texte en traitement de texte avant l'opération de segmentation) différentes « clés » qui correspondent à des variables dont on dispose par ailleurs, et qui permettent par exemple d'identifier les locuteurs, de les regrouper ensuite selon l'âge, le sexe, le niveau d'étude, etc.

Les traitements lexicométriques

Pour réaliser une segmentation automatique du texte en occurrences de formes graphiques, il suffit de choisir dans la police de caractères un sous-ensemble que l'on désignera sous le nom de *caractères délimiteurs*. Les autres caractères contenus dans la police seront considérés comme caractères *non délimiteurs*.

Une suite de caractères non délimiteurs bornée à ses deux extrémités par des caractères délimiteurs est une *occurrence*. Deux suites identiques de caractères non délimiteurs constituent deux occurrences d'une même *forme*. La forme est un archétype correspondant à un ensemble d'occurrences identiques. L'ensemble des formes d'un texte constitue son *vocabulaire*.

¹ Pour lemmatiser le vocabulaire d'un texte écrit en français, on ramène en général : les formes verbales à l'infinitif, les substantifs au singulier, les adjectifs au masculin singulier, les formes élidées à la forme sans élision.

La segmentation en formes graphiques ainsi définie nous permet maintenant de considérer le texte comme une suite d'occurrences séparées entre elles par un ou plusieurs caractères délimiteurs.

On regroupe sous le terme de *lexicométrie* toute une série de méthodes qui permettent d'opérer, à partir d'une segmentation, des réorganisations formelles de la séquence textuelle et des analyses statistiques portant sur le vocabulaire.

Ces méthodes peuvent être grossièrement regroupées de la manière suivante :

- les méthodes documentaires qui opèrent une simple réorganisation de la surface textuelle ;
- les méthodes qui opèrent, pour chaque texte pris isolément, des comptages et des calculs d'indices statistiques ;
- les méthodes statistiques « contrastives » qui produisent des résultats portant sur le vocabulaire de chacun des textes par rapport à l'ensemble des textes réunis dans un même corpus à des fins de comparaison.

Les programmes lexicométriques fournissent après la segmentation du texte en unités graphiques toute une série de documents qui permettent de mieux appréhender le vocabulaire du corpus.

L'index alphabétique permet de vérifier la saisie du texte, de rapprocher les utilisations du singulier et du pluriel d'un même substantif, les différentes flexions d'un verbe, etc.

L'index hiérarchique, dans lequel les formes sont classées par fréquence décroissante, permet d'examiner les formes les plus utilisées.

Les concordances permettent, pour chaque forme, de rassembler l'ensemble des contextes dans lesquels la forme apparaît.

Les inventaires de segments répétés permettent de repérer les séquences de formes qui apparaissent à plusieurs endroits du texte.

Le calcul des spécificités permet de dégager les formes et les segments qui se trouvent être particulièrement employés (ou, au contraire particulièrement sous-employés) par chacune des parties du corpus (LAFON, 1984).

DEUX EXEMPLES : RECHERCHE SUR DES TITRES DE THÈSES ET RECHERCHE SUR DES ENTRETIENS

Les deux exemples que nous présentons ici succinctement illustrent quelques utilisations que l'on peut faire de ce logiciel en analyse de discours sur des corpus étendus :

- dresser un premier bilan des formes utilisées ;

- orienter les hypothèses de recherche à partir des spécificités de vocabulaire repérées ;
- rechercher et éditer systématiquement dans leur contexte les formes choisies ;
- aboutir à des hypothèses motivées sur les contraintes qui s'exercent sur le discours.

Les titres de thèses

Une première recherche a été menée sur un corpus récent de 6 000 titres de thèses sur le monde arabo-musulman, puis sur un corpus (1980-1990) de 600 titres de thèses sur les villes dans les pays en développement².

Les titres de thèses se présentent généralement de manière très régulière. Ce sont des phrases nominales (très peu de verbes et pratiquement jamais conjugués) sous la forme suivante : une modalité d'énoncé (occurrence non obligatoire telle que « étude de », « recherche sur », etc.), un syntagme thématique proprement dit, pouvant être doublé d'un deuxième syntagme thématique. Dans ce corpus figurent majoritairement des localisations spatiales, plus rarement des localisations temporelles. Les titres y ont une longueur moyenne d'une douzaine de mots, mais des titres beaucoup plus longs ou beaucoup plus courts sont possibles³.

Il s'agit donc d'un *type discursif* très contraint dans sa forme, et l'on peut supposer que des contraintes fortes agissent également sur les thématiques et sur la construction des objets (contraintes disciplinaires notamment). Collecter ces titres pour en constituer un corpus est donc un choix opéré⁴ par le documentaliste ou le chercheur, à partir de l'hypothèse que la réunion de ces énoncés, tous issus d'énonciations singulières, est significative d'un ensemble discursif pertinent. Le titre de thèse renvoie à la fois à un contenu (le texte de la thèse elle-même) et à un champ discursif (l'espace de la discipline)⁵.

² Corpus de thèses soutenues dans les universités françaises, réunis respectivement par les documentalistes de l'Iremam et par Laurent Vidal (SANTO MARTINO et LEIMDORFER, à paraître ; VIDAL et LEIMDORFER, 1992).

³ Le titre le plus court rencontré dans ce corpus est d'un seul mot, qui porte, il est vrai, plusieurs informations : « Iranité ». Les titres les plus longs peuvent comporter jusqu'à 30 mots environ, combinant et comparant plusieurs thématiques. Exemple : « Incertitudes, rigueur et arbitraire dans la planification spatiale des grandes régions urbaines. Deux études de cas : la reconstruction du Grand Beyrouth et l'aménagement de la région Ile-de-France. »

⁴ À partir de certains critères qui sont autant d'hypothèses sur les domaines discursifs et pratiques (dans la note 3 « l'Orientalisme » et les recherches « urbaines »).

⁵ Sur ces questions, voir les deux articles cités dans la note 3.

À chaque titre correspond un certain nombre de données : année de soutenance, région géographique, discipline de soutenance, ville universitaire. Ces données sont indiquées dans le corpus sous forme de « clés » qui le divisent en « parties ». Le logiciel permet dès lors de savoir si le vocabulaire se répartit de manière spécifique selon ces variables. Dans ce cas précis, il se trouve que les disciplines sont le mode de partition le plus significatif (la période d'étude est trop courte, une quinzaine d'années, pour dégager des tendances claires). La répartition du vocabulaire se dégage assez nettement par discipline, mais les termes jugés « banals » sont également intéressants par la connaissance qu'ils nous donnent du vocabulaire « partagé ». Voici la répartition entre spécificités positives et négatives entre disciplines dans le corpus de thèses « urbaines - pays en développement » (VIDAL et LEIMDORFER, 1992 : 26-29) :

Spécificités positives

organisation (spatiale, urbaine)
 agglomération
 croissance (urbaine)
 aménagement
 équipements
 espace
 petites villes
 ville moyenne

transports (collectifs)
 politique
 planification
 techniques
 logement

sociale
 pratique
 vie
 milieu (urbain)
 populaire
 moderne
 structure

siècle
 histoire
 époque

informel (secteur)
 pays (en développement)
 urbaines
 économie
 analyse
 politiques
 emploi

*Géographie**Urbanisme**Sociologie**Histoire**Économie***Spécificités négatives**

analyse
 logement
 informel (secteur)
 milieu (urbain)
 pratique
 transport(s)

espace (urbain)
 croissance (urbaine)

agglomération
 région
 aménagement
 organisation

cas
 urbaine

ville

Spécificités positives

migrations
transport
logements

droit
urbanisme
aménagement

Spécificités négatives

Économie (suite)

Droit

Les termes entre parenthèses désignent des « segments répétés » : organisation, organisation spatiale, organisation urbaine ont tous trois des spécificités positives. Rappelons que les « spécificités » sont des calculs de probabilité d'apparition d'un mot dans une partie d'un texte. Une spécificité négative n'implique pas l'inexistence d'un mot dans une partie, mais une « sous-représentation » de ce mot dans cette partie.

La géographie, discipline centrale des études urbaines, et qui en tant que telle exerce un poids important (la moitié des thèses), s'oppose aux autres, selon diverses modalités :

— à la sociologie sur les mots « agglomération », terme technique (le mot « ville » est banal, sauf en économie), « organisation » (« spatiale », « urbaine »), « aménagement » (que par contre la géographie partage avec le droit). Sociologie et géographie s'opposent nettement aussi sur « milieu », « vie », « populaires », « social », « rapports », « pratique », qui cette fois sont des termes techniques propres à la sociologie :

— à l'économie sur le mot « informel » (« secteur informel »), là encore terme technique propre à cette discipline, et sur « analyse » ;

— plus délicats sont les rapports de la géographie à l'urbanisme. L'urbanisme se refuse aux études sur l'« espace » (y compris urbain), et se concentre sur des « secteurs » particuliers : « transports », « logement », « eau », par des études de cas (qui à leur tour ne font pas partie de l'arsenal méthodologique de l'histoire). De même, ce qui est de l'ordre de l'action des pouvoirs est concentré dans les études d'urbanisme : « politique », « planification ». Plus curieusement la « croissance » (et la « croissance urbaine ») n'entre pas dans son champ. Tout se passe comme si géographie et urbanisme, dont on peut penser qu'elles sont des disciplines concurrentes, s'étaient réparties les domaines : les études du développement urbain dans le monde arabo-musulman et l'Afrique pour la géographie, les études sectorielles sur l'action des pouvoirs publics dans les pays latino-américains pour l'urbanisme. Il y a là peut-être aussi des lignes de force institutionnelles qui organisent les champs disciplinaires.

L'échelle d'observation est aussi un élément de différenciation important : le pays et non la ville pour l'économie, le milieu pour la sociologie et non la région.

Les « lieux communs » des disciplines dans le domaine urbain sont avant tout les termes processuels (« évolution », « processus », « dynamique », « transformations »). Une étude plus fine serait nécessaire pour départager les mots qui renvoient à une utilisation « naturelle » (« quartier », « commerce », « activités », « population », « banlieue ») et ceux caractéristiques d'un emploi technique, mais partagé, dans le champ de l'urbain (« urbain », « habitat », « urbanisation », « rapports ville-campagne », « périphérie »).

Par ailleurs, le logiciel permet une comparaison entre corpus différents, ce qui donne des indications parfois éclairantes. En réunissant deux corpus (de conditions de production proches) dans un même texte, on a pu comparer titres de thèses et titres de rapports d'études dans le domaine urbain (chacun des corpus formant dès lors une « partie » de ce texte). Les différences de l'énonciation entre ces deux types discursifs y sont bien marquées : d'un côté, des objets globaux, des catégories (les thèses : une construction d'objet), de l'autre, des classes d'objets à comparer (les rapports d'études : une énonciation « d'expertise ») ; pour les thèses un vocabulaire conceptuel global (« urbanisation », « croissance », etc.), pour les rapports d'études des termes « techniques » (« ped », « gestion », etc.). De même, la mise en relation des thèses « modernes » et des thèses de l'époque coloniale sur l'Algérie (LEIMDORFER, 1992), indiquait les ruptures de thèmes (disparition du « droit musulman », importance dans les titres de la mise en relation de phénomènes dans les thèses modernes).

Les entretiens

La deuxième recherche concernait trois interviews de patron et patronnes de « maquis » (restaurants populaires à Abidjan). La situation interlocutoire et l'énonciation sont bien entendu très différentes de celles de l'exemple précédent. Il s'agit d'entretiens oraux, en situation de face à face, où le discours s'organise autour des questions, des reprises, de la communauté discursive créée par la relation. Le discours produit est — contrairement au titres de thèses — un discours provoqué, et retranscrit⁶.

Les tris se sont faits sur les locuteurs : les interviewers et les interviewé(e)s, d'une part, et entre les interviews, d'autre part. Les résultats

⁶ L'activité du chercheur est dans un cas une activité de collecte et de constitution de corpus, dans l'autre s'y ajoute la création d'une situation interlocutoire particulière et la transformation d'une donnée situationnelle orale en transcription.

permettent de voir clairement les différences de langage entre les locuteurs, et notamment de mettre à plat la part des interviewers dans l'interrelation, mais aussi les différences entre les interviewé(e)s.

L'analyse qui avait été faite avant l'utilisation du logiciel avait repéré trois placements identitaires différents : une interviewée se voyait « restauratrice » et pensait en terme de statut ; une autre s'identifiait à sa « famille » et au rôle qu'elle y jouait dans la trame des solidarités ; un troisième se voyait « commerçant », pour lequel il n'y avait pas de différences fondamentales entre « maquis » et « restaurant ». Les tris lexicométriques opérés ont confirmé ce point de vue : spécificités très fortes des termes « restaurant », « famille », « commerce » dans les interviews respectives.

De plus, les concordances permettent aisément une comparaison des marques de l'énonciation personnelle : ainsi les rapports entre « je », « ils/elles », « nous » et, en particulier, la construction de groupes inclusifs et exclusifs (l'opposition entre « nous » et « eux/ils/elles »).

Cela étant, il faut tenir compte de la forme particulière du discours oral et de sa linéarité : ainsi les répétitions de mots jouent fréquemment à l'oral un rôle de reprise thématique (maintien du fil du discours) ou d'insistance. Dans la mesure où le logiciel fait des comptages quantitatifs, il est donc indispensable de réfléchir au statut de certaines occurrences répétitives. Les résultats montrent de même la prépondérance des marques d'énonciation verbale et des marques de l'interlocution : pronoms personnels, « être », « avoir », démonstratifs, « oui », « non », des présentatifs, « c'est », etc.

Les tableaux en annexe présentent des exemples de *concordances* et de calcul de *spécificités* pour les deux exemples évoqués. Nous avons choisi les termes « problème » et « Afrique », qui sont communs aux deux corpus, et qui illustrent en quoi l'emploi et la construction du sens diffèrent profondément lorsqu'on les analyse dans leur contexte⁷. Dans l'entretien, « Afrique » choisi comme terme pivot (avec un classement alphabétique du contexte droit), sert de marque identitaire et entraîne des éclaircissements culturels destinés sans doute en partie à l'interviewer et en partie à synthétiser une situation. Dans le corpus des titres, « Afrique » permet de localiser et de circonscrire un champ d'étude (tri du contexte gauche⁸). « Problème » est la plupart du temps dans l'en-

⁷ Les contextes peuvent être classés dans l'ordre du texte ou alphabétiquement à gauche (avant le mot pivot) ou à droite (après). Il est par exemple plus intéressant de classer à gauche les contextes d'un adjectif (majoritairement placés en français à droite du substantif). Une erreur dans le choix de tri n'empêche cependant pas la comparaison, dans la mesure où l'ensemble des contextes (jusqu'à trois lignes) est restitué.

⁸ « Afrique » pour les entretiens « ouvre » sur une prédication, alors qu'il intervient pour les titres comme localisation en fin de phrase.

retien la condensation d'un événement et d'une situation, attribué à un ou des locuteurs (« avoir un problème »), alors que dans les titres, il est spécifié par un syntagme nominal (« problème d'urbanisme, de logement », etc.⁹). L'analyse des deux corpus se fait donc dans des directions différentes : construction des objets et signification dans un champ (par exemple les occurrences et les contextes du mot « ville ») pour les titres de thèses, énonciation, normes et placements identitaires (occurrences et oppositions entre les « je », les « nous », les « ils/elles » par exemple) pour les entretiens.

ÉLÉMENTS DE CONCLUSION

L'outil lexicométrique s'avère, du point de vue de l'analyse de discours, d'un très grand intérêt, dans trois directions principales :

- par les données quantitatives fournies, les comparaisons et les vérifications qu'il permet ;
- comme outil de repérage de pistes de recherche, et comme premier bilan d'un corpus ;
- comme outil heuristique puissant, entraînant à des allers-retours fructueux entre le texte analysé et les données produites. Il incite à une définition plus fine des données et à des comparaisons vers d'autres corpus. Il oblige également à une réflexion sur le statut du « quantitatif » dans le discours à l'écrit et à l'oral.

D'un point de vue pratique, il est particulièrement utile si plusieurs éléments se trouvent réunis :

- si le corpus est relativement important, difficilement maîtrisable par une analyse fine de fragments ; mais des informations intéressantes se dégagent avec des corpus de quelques dizaines de pages seulement ;
- si la saisie sous traitement de texte peut être faite sans difficultés particulières, et de manière économique (en temps de travail notamment) ;
- si le corpus est suffisamment connu, déjà analysé pour que les indications statistiques données puissent prendre sens et orienter la recherche ; lorsque cette connaissance de « l'intérieur » n'existe pas, les données fournies par le calcul indiquent autant de pistes possibles, mais qu'il est difficile d'examiner exhaustivement. Par contre, lorsque le corpus a déjà été analysé en partie ou que l'on dispose de pistes de recherche identifiées, l'outil lexicométrique devient un « multiplica-

⁹ Donc avec un tri à droite pour les titres et un tri à gauche pour les entretiens.

teur » de recherche remarquable, par les allers-retours continuels qu'il permet entre l'analyse de fragments, les données statistiques et les nouvelles demandes de tri que l'on peut formuler.

BIBLIOGRAPHIE

- ACHARD (P.), 1991. — *Une approche discursive des questionnaires : l'exemple d'une enquête pendant la guerre d'Algérie, Langage et Société*, n° 55, mars 1991 : 5-40.
- LAFON (P.), 1984. — *Dépouillements et statistiques en lexicométrie*, Genève-Paris, Slatkine-Champion.
- LEBART (A.), SALEM (A.), 1994. — *Statistiques textuelles*, Dunod, 342 p.
- LEIMDORFER (F.), 1992. — *Discours académique et colonisation*, Publisud, 330 p.
- SALEM (A.), 1987. — *Pratique des segments répétés*, Paris, Klincksieck, INDLP, 333 p.
- SANTO MARTINO (R.), LEIMDORFER (F.). — Figures d'une construction disciplinaire, à propos des études urbaines dans le champ de la recherche doctorale sur l'aire arabomusulmane, à paraître dans les *Cahiers d'URBAMA*, 1994.
- VIDAL (L.), LEIMDORFER (F.), 1992. — Les thèses françaises sur les villes des pays en développement (1980-1990), *Pratiques Urbaines, Interurba*, 159 p.

ANNEXE

Exemples de concordances et de spécificités

Concordances

--- Requête num 1 (Entretiens) --- afrrique -> F= 9
 --- Concordance --- tri après

L 1242 plus + c' est avant on dit en afrrique + mon oncle est là- bas je vais
 L 651 tant les enfants NADU- ouli en afrrique c' est ça + c' est pourquoi je
 L 999 i + quand tu es sans enfant en afrrique c' est très dur + tu es quand t
 L 999 de contre lui + nous sommes en afrrique ça se dit pas ça + dire que une
 L 1316 t à fait normal nous sommes en afrrique elle a qu' à partir + FRANÇOIS-
 L 995 ur les enfants des autres + en afrrique heu c' est + c' est nécessaire
 L 995 + oul + parce que chez nous en afrrique heu tu peux demander partout si
 L 1106 ANTOINETTE- pourquoi? NADU- en afrrique lui tout ceux qui ont adopté en
 L 1110 ein + donc heu une personne en afrrique sans enfant c' est pas facile +

--- Requête num 2 --- (Thèmes) afrrique -> F= 12
 --- Concordance --- tri avant

L 617 ismes d' une ville capitale d' afrrique centrale: kinshasa. problème d'
 L 21 n- baptiste gramaye, évêque d' afrrique l' organisation urbaine des bas
 L 513 et citoyenneté d' une ville d' afrrique centrale, bangassou et sa régio
 L 549 es et processus migratoires en afrrique de l' ouest. le cas de trois ce
 L 545 banisation sur les paysages en afrrique et en chine. secteur informel e
 L 305 aspect du changement social en afrrique: l' attrait de la ville de haaa
 L 9 isation de l' espace urbain en afrrique tropicale. l' urbanisation, un
 L 11 duction de l' espace urbain en afrrique centrale, droit et ville en afr
 L 6 as de kpon émeutes urbaines en afrrique, l' ensablement des villes oua
 L 4 ansports collectifs urbains en afrrique noire francophone(...). l' évol
 L 183 a relations campagne- ville en afrrique centrale, au- delà du dualisme:
 L 12 ue centrale, droit et ville en afrrique noire. essai sur le droit de l'

--- Requête num 12 --- (Entretiens) problème -> F= 27
 --- Concordance --- tri avant

L 1232 ussis + je réussis à régler ce problème FRANÇOIS- et ce que tu as dit
 L 1384 e fois que je vais résoudre ce problème là je réussis à résoudre FRANÇ
 L 1272 te je lui ai dit quel genre de problème? + il dit beh oulà y a une ap
 L 876 veux les rejoindre y a pas de problème + j' ai même perdu leur adress
 L 1086 rrait me jeter comme ça? + mon problème c' est ça + et comment il a fa
 L 1034 ai fini par lui expliquer mon problème et ça lui a fait très mal + et
 L 1126 as + en fait + je n' avais pas problème de ventre je n' avais pas aux
 L 1272 dis y a un problème où et quel problème? + il dit sur votre terrain là
 L 1384 respecte il vient confier son problème + et chaque fois que je vais r
 L 1224 premier secrétaire a hum a un problème avec sa femme hum + sa il a ép
 L 1272 venu vous voir + je dis y a un problème où et quel problème? + il dit
 L 327 nger les affaires quand y a un problème NADU- oul + jamais je fais pal
 L 1384 que une femme heu quand y a un problème quand les gens ont un problè
 L 1272 dit pour rien parce que y a un problème c' est pourquoi je suis venu v
 L 1224 écart parce que l' autre à un problème avec sa femme + le + premier s
 L 1034 me? je dis comment ça j' ai un problème pourquoi j' ai un problème + t
 L 1034 un problème pourquoi j' ai un problème + tu crois que j' ai un problè
 L 1034 oblème + tu crois que j' ai un problème + il dit si si + et puis moi q
 L 1034 bas + puis il a dit + tu as un problème? je dis comment ça j' ai un pr
 L 1126 tu sais) + il dit non tu as un problème pourquoi est- ce que tu veux p
 L 1034 it je vois en toi que tu as un problème et que tu me caches + je dis m
 L 1021 j peut définir ça + y avait un problème là dans mon ventre + et quand
 L 125 NÇOIS- parce que il y avait un problème d' autoroute je crois NADU- ou
 L 1374 ne à qui quand vous confiez un problème ça se règle facilement NADU- f
 L 1384 problème quand les gens ont un problème ils viennent me voir + et puis
 L 1232 e quand on m' envoie régler un problème + je réussis + je réussis à ré
 L 1272 a + quand il m' a dit que y a un problème toute suite je lui ai dit quel

=== Requête num 1) === (Thèses) problèmes -> F= 12
 --- Concordance --- tri après

L 618 d'afrique centrale: kinshasa. problème d'approvisionnement des ville
 L 219 e moyenne égyptienne: rosette. problème d'urbanisme et d'aménagement
 L 258 sance urbaine en irak. bagdad, problème de croissance urbaine. les mon
 L 107 nvasions, réponse politique au problème de l'habitat populaire. analy
 L 23 conomie planifiée. l' étude du problème de l'insertion des adolescent
 L 237 ires de libreville(gabon). le problème de la communication chez les t
 L 598 nstruction: une alternative au problème de la concentration urbaine au
 L 268 ir du vieux shiraz: analyse et problème de réhabilitation des quartier
 L 590 os. la réponse des citoyens au problème du logement dans la ville indu
 L 93 résidentielle au tertaire. le problème du logement urbain au brésil.
 L 606 terres urbaines à caracas: un problème ou une solution? amin la asmo
 L 342 historique le droit de cité le problème posé par la réhabilitation de

Spécificités du corpus des titres de thèses urbaines

-- Partie numéro 1 -- code : 1 (géographie)--

n°	terme	F	f	spec.	orig.
25	croissance	47	34	+E04	
31	organisation	26	22	+E04	
513	petites villes	7	7	+E03	*
109	périphérique	10	9	+E03	
13	urbaine	119	71	+E03	
325	l organisation	15	13	+E03	
77	a	13	11	+E03	
41	agglomération	21	16	+E03	
370	dans l organisation	7	7	+E03	*
463	organisation de l	10	9	+E03	
305	l agglomération	18	15	+E03	
406	urbaine et	26	19	+E03	
82	république	13	11	+E03	
238	de l agglomération	5	5	+E02	*
174	équipements	6	6	+E02	*
217	port	5	5	+E02	*
466	organisation urbaine	5	5	+E02	*
224	sidi	5	5	+E02	*
34	aménagement	26	18	+E02	
103	constantine	10	8	+E02	
311	l espace	56	35	+E02	
514	sa région	5	5	+E02	*
141	métropole	8	7	+E02	
367	dans l agglomération	10	8	+E02	
495	rôle dans	6	6	+E02	*
360	du congo	8	7	+E02	
107	mexique	10	8	+E02	
64	médina	16	12	+E02	
510	fonciera et	5	5	+E02	*
176	géographique	6	6	+E02	*
234	de croissance	5	5	+E02	*
66	actjvités	15	11	+E02	
261	et aménagement	5	5	+E02	*
110	petites	10	8	+E02	
369	dans l organisation de l	5	5	+E02	*
142	moyenne	8	7	+E02	
287	la médina de	8	7	+E02	
198	commerciales	5	5	+E02	*
112	sa	10	8	+E02	
117	structures	10	8	+E02	
288	la médina	15	12	+E02	
464	organisation de l espace	9	8	+E02	
98	péri	11	9	+E02	
19	espace	69	42	+E02	
451	croissance urbaine	29	20	+E02	
371	dans l	35	24	+E02	
352	à l époque	5	0	-E02	
515	histoire de	5	0	-E02	
143	rapports	8	1	-E02	
319	l exemple de la	7	0	-E02	
139	histoire	8	1	-E02	
105	informel	10	1	-E02	
116	social	10	1	-E02	
460	politique de l	5	0	-E02	
222	ségrégation	5	0	-E02	
208	historique	5	0	-E02	
216	phénomène	5	0	-E02	
497	secteur informel	8	1	-E02	
93	vie	12	2	-E02	
473	exemple de	14	3	-E02	
223	sfax	5	0	-E02	

474	urbaine en	5	0	-E02
166	application	6	0	-E02
22	sur	55	19	-E02
189	transport	6	0	-E02
472	exemple de la	8	1	-E02
354	à l	29	8	-E02
398	des transports	8	1	-E02
205	époque	5	0	-E02
320	l exemple de	13	2	-E02
94	brésil	11	2	-E02
377	dans les	25	6	-E02
29	analyse	31	8	-E03
100	société	11	1	-E03
32	logement	27	6	-E03
55	sociale	18	2	-E03
420	en milieu urbain	12	0	-E04
419	en milieu	13	0	-E04
65	milieu	16	1	-E04
499	milieu urbain	14	0	-E05

Fin partie: 1

formes sp+ : 24 sp- : 19
segments sp+ : 22 sp- : 15

-- Partie numéro 2 -- code : 2 (urbanisme) --

n°	terme	F	f	spec.	orig.
108	paulo	10	8	+E05	
460	politique de l	5	5	+E04	*
30	politique	30	14	+E04	
69	eau	15	8	+E03	
291	la politique	16	8	+E03	
310	l eau	8	6	+E03	
518	são paulo	7	5	+E03	
191	aire	5	4	+E03	
106	métropolitaine	10	6	+E03	
206	évaluation	5	4	+E03	
162	são	7	5	+E03	
398	des transports	8	6	+E03	
306	l aire métropolitaine	5	4	+E03	
459	politique de	6	5	+E03	
181	leurs	6	4	+E02	
268	et l	18	8	+E02	
5	d	212	54	+E02	
503	eau et	6	4	+E02	
14	ese	104	30	+E02	
47	transports	20	8	+E02	
256	de rabat sole	7	4	+E02	
32	logement	27	10	+E02	
297	la région	14	7	+E02	
299	la ville d	6	4	+E02	
511	métropolitaine de	8	5	+E02	
484	processus de	6	4	+E02	
59	planification	17	8	+E02	
397	des transports collectifs	6	4	+E02	
432	espace urbain	23	1	-E02	
311	l espace	56	5	-E02	
23	maroc	51	4	-E02	
19	espace	69	7	-E02	
18	urbain	84	10	-E02	
442	urbanisation et	14	0	-E02	
451	croissance urbaine	29	2	-E02	
25	croissance	47	3	-E03	

Fin partie: 2

formes sp+ : 13 sp- : 4
segments sp+ : 15 sp- : 4