Une plate-forme logicielle dédiée à la cartographie thématique de corpus

Thibault ROY
Laboratoire GREYC, Equipe ISLanD - Université de Caen / Basse-Normandie
Campus II - Côte de Nacre - Bd Maréchal Juin - 14032 Caen Cedex
troy@info.unicaen.fr
date de soutenance prévue fin 2007

Mots-clefs – Keywords

cartographie de corpus, analyse thématique, logiciel individu-centré, analyse des données textuelles
corpora cartography, thematic analysis, user-centered software, textual data analysis

Résumé - Abstract

Cet article présente les principes de fonctionnement et les intérêts d'une plate-forme logicielle centrée sur un utilisateur ou un groupe d'utilisateurs et dédiée à la visualisation de propriétés thématiques d'ensembles de documents électroniques. Cette plate-forme, appelée ProxiDocs, permet de dresser des représentations graphiques (des cartes) d'un ensemble de textes à partir de thèmes choisis et définis par un utilisateur ou un groupe d'utilisateurs. Ces cartes sont interactives et permettent de visualiser les proximités et les différences thématiques entre textes composant le corpus étudié. Selon le type d'analyse souhaitée par l'utilisateur, ces cartes peuvent également s'animer afin de représenter les changements thématiques d'un ensemble de textes au fil du temps.

This article presents a user-centered software dedicated to the visualization of thematic properties of sets of electronic documents. This software, called ProxiDocs, allows its users to realize thematic maps from a corpora and themes they choose and defined. These maps are interactive and reveal thematic proximities and differences between texts composing the studied corpus. According to the analysis wished by the user, maps can be animated in order to represent thematic changes of the analysed set of texts relating to the time.

1 Introduction

Cet article présente les principes de fonctionnement et les intérêts de la plate-forme logicielle ProxiDocs dédiée à des analyses thématiques de corpus de textes. Sur le modèle de (Pichon et Sébillot, 1999), nous entendons par *thèmes*, les sujets abordés dans un texte. Traiter la thématique d'un texte revient donc pour nous à mettre en évidence les principaux sujets abordés dans ce dernier. Dans un grand nombre de situations (extraction et recherche d'information, analyse de flux documentaires, etc.), l'appréhension des thèmes abordés dans des textes constitue une première analyse importante et délicate.

ProxiDocs a pour objectif d'aider ses utilisateurs dans de telles situations en leur fournissant des représentations graphiques (que nous appelons des *cartes thématiques*) d'un corpus de textes donné. Les cartes construites mettent en évidence la répartition des thèmes au sein des textes du corpus et révèlent des proximités et des différences de thèmes entre textes. Ces cartes sont construites à partir de thèmes choisis et définis par l'utilisateur en fonction de la tâche qu'il souhaite accomplir. En ce sens, c'est un système *anthropocentré* tel que le définit (Thlivitis, 1998) : son exécution n'est pas guidée par des ressources propres, les traitements réalisés sont personnalisés et intégralement conditionnés par les besoins et les choix de l'utilisateur.

La plate-forme ProxiDocs est *open-source*, développée en Java et disponible avec sa documentation sur le Web[1]. C'est un système développé à la façon d'un logiciel d'étude au sens de (Nicolle, 1996), c'est-à-dire qu'il est conçu dans le but de vérifier des hypothèses sur les langues en les expérimentant sur du matériau textuel attesté. ProxiDocs fait partie d'un ensemble de logiciels d'étude en constante évolution dédiée à l'analyse linguistique informatisée de corpus de documents électroniques[2] développés au sein de l'équipe ISLanD du laboratoire GREYC.

Dans cet article, nous présentons tout d'abord des outils utilisant des techniques de cartographie afin d'accéder aux informations contenues dans des collections de documents. Ensuite, nous abordons tout particulièrement la plate-forme ProxiDocs en détaillant ses principes de fonctionnement et en présentant les différentes types de cartes qu'elle permet de construire.

2 La cartographie d'ensembles de documents textuels

Le nombre de documents textuels produits et échangés chaque jour ne cesse de croître. Afin d'isoler les principales informations contenues dans des ensembles de documents textuels, il peut être intéressant de proposer des représentations graphiques de ces ensembles. De telles représentations permettent de faire intervenir une notion de proximité entre éléments. Selon le type d'analyse réalisée, il est alors possible d'observer des similarités et des différences de styles, de thèmes, de mises en forme entre documents d'un même ensemble.

Depuis quelques années, des outils d'analyse textuelle exploitent une technique de visualisation appelée *cartographie*. À la manière d'une carte routière mettant en évidence des villes et des routes les reliant, une carte d'un ensemble de données textuelles met en évidence des proximités et des liens entre entités textuelles, tels des mots, des textes, etc.

Dans une tâche d'extraction d'information, les auteurs de (Mokrane et al., 2004) propose d'utiliser une technique de cartographie afin de visualiser les liens entre les principaux termes présents

[1] http://www.info.unicaen.fr/~troy/proxidocs

[2] http://www.greyc.unicaen.fr/island/logiciel

dans un ensemble de dépêches d'agences de presse.

Depuis 2001, les deux métamoteurs de recherche cartographiques KartOO (Chung et al., 2002) et MapStan (Spinat, 2002) sont disponibles sur le Web[3]. En réponse à une requête de l'utilisateur, ces deux outils retournent des cartes représentant les sites proposés en réponse à cette requête. Les sites jugés similaires par le système sont alors situés à proximité sur les cartes et il est ainsi possible de distinguer les grandes catégories d'informations proposées en réponse à la requête de l'utilisateur.

Pour une tâche de parcours rapide d'un ensemble documentaire, le logiciel NeuroNav (Lelu et Aubin, 2001) de la société Diatopie[4] présente sur une carte des groupes de documents. Les différents groupes déterminés par le système et la disposition de ces groupes sur la carte peuvent ainsi indiquer des proximités de contenu entre documents et groupes représentés.

De nombreux logiciels dédiés à l'analyse de données textuelles proposent également des résultats d'analyses sous forme de cartes. Parmi ces logiciels, nous pouvons citer Hyperbase d'Etienne Brunet[5], Lexico3 de l'équipe CLA2T de Paris III[6] ou encore Lexica de la société Le Sphinx[7].

À la manière de la plupart des outils précédents, la plate-forme que nous proposons va utiliser une technique de cartographie afin de mettre en évidence des proximités et des liens entre textes d'un même ensemble. Contrairement à ces outils, les traitements réalisés par ProxiDocs prennent en considération les particularités de l'utilisateur et de sa tâche. À partir de ressources décrivant le point de vue de l'utilisateur sur des thèmes de son choix, ProxiDocs cherche à extraire les tendances thématiques des textes d'un corpus. Ces tendances sont ensuite mises en évidence sur des cartes.

3 La plate-forme logicielle ProxiDocs

3.1 Définitions et intérêts

La plate-forme ProxiDocs permet de construire différents types de cartes thématiques à partir d'un corpus de textes et de thèmes choisis et définis par l'utilisateur :

- des cartes en 2 ou 3 dimensions représentant chaque texte du corpus analysé par un point, les proximités entre points indiquent des similarités de thèmes abordés dans les textes représentés (figure 2) ;
- des cartes en 2 ou 3 dimensions mettant en évidence des groupes de textes abordant des thèmes proches, chaque groupe est alors représenté sur la carte par un disque ou une sphère de diamètre proportionnelle à nombre de textes qu'il contient (figure 3) ;
- et des cartes en 2 dimensions animées mettant en évidence l'évolution des thèmes abordés dans les textes du corpus au fil du temps (figure 4).

Les intérêts de ProxiDocs sont multiples. Les cartes thématiques construites par la plate-forme permettent d'observer des regroupements entre textes abordant des thèmes proches. De tels

[3] Respectivement disponibles aux adresses : `http://www.kartoo.fr` et `http://www.mapstan.net`
[4] `http://www.diatopie.com/`
[5] `http://ancilla.unice.fr/~brunet/pub/hyperbase.html`
[6] `http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicowww/lexico3.htm`
[7] `http://www.lesphinx-developpement.fr/`

regroupements sont très utiles dans des tâches de parcours rapide d'un grand ensemble de textes. Dans des tâches de recherche documentaire sur Internet, la cartographie thématique appliquée aux pages désignées par un moteur de recherche en réponse à une requête permet d'observer des regroupements entre pages abordant des thèmes proches. De cette manière, l'utilisateur peut orienter sa recherche selon les thèmes l'intéressant dans le cadre de sa recherche. Dans le cadre de la veille documentaire, la cartographie d'un ensemble de documents sur différentes périodes consécutives illustre des changements de thèmes au fil du temps.

3.2 Les thèmes utilisés

Afin de construire des cartes thématiques d'un corpus de textes, l'utilisateur doit tout d'abord choisir et définir les thèmes qu'ils souhaitent faire intervenir sur ses cartes. Chaque thème est représenté par une liste de lexies (mots simples ou mots composés) lui étant associées du point de vue de l'utilisateur. Afin de simplifier la phase de construction des thèmes, l'utilisateur n'indique que les formes lemmatisées des lexies. La plate-forme intégre une étape de génération de formes fléchies à l'aide d'une base de données lexicales. Un utilisateur peut par exemple associer les lexies suivantes au thème de l'aviation : avion, appareil, vol, pilote, pilotage, piloter, passager, Boeing, Air France, décollage, etc.

Deux logiciels sont proposés à l'utilisateur afin de l'aider à construire ses thèmes :
- l'outil Memlabor (Perlerin, 2002) permettant une analyse statistique des graphies répétées d'un corpus. En exploitant le principe de cohésion lexicale, MemLabor se fonde sur l'hypothèse que plus une graphie (hors mots d'un anti-dictionnaire contenant par exemple les mots grammaticaux) est répétée dans le corpus, plus elle est susceptible de pouvoir être associée à l'un des thèmes présents dans le corpus (Perlerin, 2004, p. 141). En présentant à l'utilisateur une liste des graphies classées par ordre décroissant de fréquence d'apparition, le logiciel permet une première assistance à l'extraction de graphies intéressantes pour l'utilisateur selon sa tâche à partir d'un corpus.
- l'outil ThemeEditor (Beust, 2002) permettant de composer des graphies en lexies et de les rassembler en thèmes. Ce rassemblement est non exclusif, une lexie pouvant être associée à plusieurs thèmes. Les ressources ainsi constituées sont projetées sur le corpus initial par une annotation XML. Un principe de surlignage avec différentes couleurs (une couleur correspondant à un thème) permet de mettre en évidence la répartition, l'alternance et les enchaînements au long d'un texte des thèmes ainsi créés.

Les thèmes construits à l'aide des logiciels précédents sont stockés en machine sous forme de listes au format XML afin de permettre une facile réutilisation.

3.3 Les traitements réalisés

La plateforme ProxiDocs prend en entrée un fichier XML contenant des thèmes construits par l'utilisateur et un corpus de documents électroniques au format texte ou HTML[8]. Les différents traitements réalisés par ProxiDocs sont présentés en figure 1. Les cartes thématiques produites en sortie de l'application sont représentées dans le format SVG (W3C, 2001), ceci afin de garan-

[8]Dans le cas des documents HTML, seules les parties textuelles sont traitées. Les informations concernant la structure du document et les éléments qu'il contient (telles les images) ne sont pas encore prises en considération dans nos analyses mais pourront l'être dans des prochaines versions de la plate-forme.

tir leur portabilité sur différents systèmes et de permettre à l'utilisateur d'effectuer facilement des zooms et des déplacements.

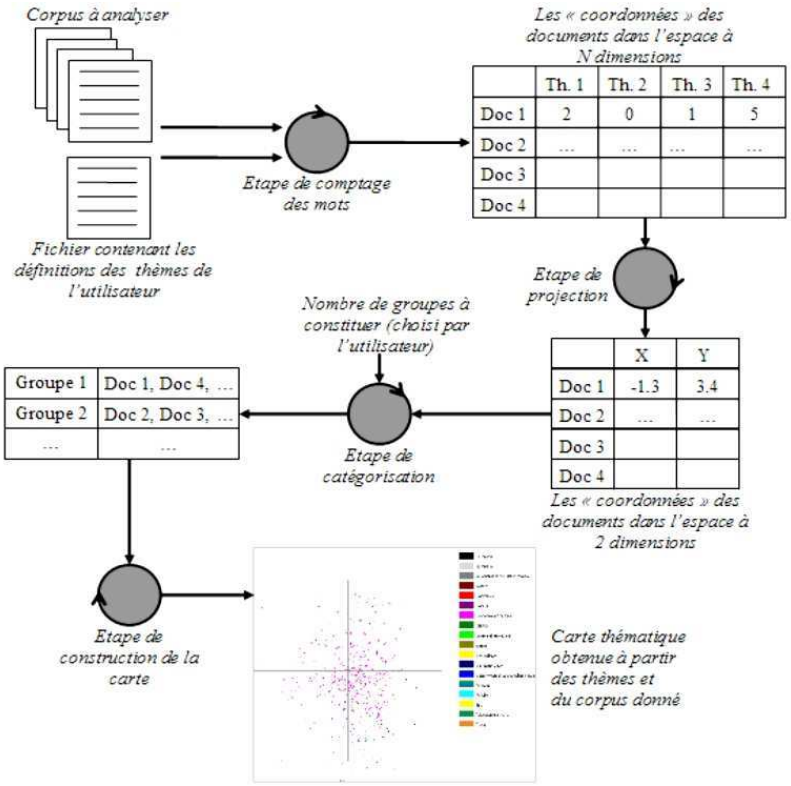

FIG. 1 – Chaîne de traitement de la plate-forme ProxiDocs

Le premier traitement réalisé par la plateforme est un comptage des occurrences de lexies de chaque thème et de leurs formes fléchies dans chaque texte. Un *vecteur* d'entiers de dimension égale au nombre de thèmes choisis et définis par l'utilisateur est alors associé à chaque texte. Supposons qu'un utilisateur fasse intervenir dans la construction de ses cartes thématiques les thèmes de la Bourse, de l'Économie, de la Météo et du Sport, les vecteurs représentant les textes sont de la forme :

Vecteur(Texte) = (nb_lexies(Bourse), nb_lexies(Économie), nb_lexies(Météo), nb_lexies(Sport))

Cette méthode de comptage est appelée *absolue*, du fait qu'elle ne fait pas intervenir la taille des textes lors du comptage. Une seconde méthode, appelée *relative*, détermine pour chaque texte du corpus, les pourcentages des lexies de chaque thème et de leurs formes fléchies par rapport au nombre total de mots du texte. Cette méthode est particulièrement intéressante lorsque la taille des textes du corpus varie significativement.

Les vecteurs obtenus à l'issue de l'étape de comptage prennent place dans des espaces de dimensions égales aux nombres de thèmes définis par l'utilisateur[9]. Pour visualiser ces vecteurs sur des cartes en 2 ou 3 dimensions, il faut réaliser une *projection* de ces vecteurs. Pour cela, nous proposons plusieurs méthodes d'analyse des données dont l'*Analyse en Composante Principales* (ACP) (Bouroche et Saporta, 1980) et l'*Analyse Factorielle des Correspondances* (Benzécri, 1980). À l'issue de cette étape de projection, nous proposons aux utilisateurs des regrou-

[9]Dans l'exemple précédent, les vecteurs représentant les textes prennent place dans un espace à 4 dimensions.

pements automatiques des textes sur les cartes. Pour cela, nous avons intégré une méthode de catégorisation, appelée *Catégorisation Hiérarchique Ascendante* (Bouroche et Saporta, 1980).

Afin de mettre en évidence les résultats produits par l'enchaînement de ces différents traitements, nous présentons dans la partie suivante des exemples de cartes thématiques construites par ProxiDocs à partir d'un corpus d'articles de presse et de thèmes que nous avons définis. Pour plus de détails sur ces traitements, nous renvoyons à (Roy et Beust, 2004).

3.4 Les cartes thématiques obtenues sur un exemple

Les cartes présentées dans cette section ont été construites à partir d'un corpus constitué de 789 articles de l'année 1989 du journal "Le Monde" totalisant environ 700 000 graphies. Le jeu de thèmes utilisé est généraliste et propose des descriptions des 18 thèmes suivants : la justice, la religion, la violence, l'éducation, l'agriculture, la sécurité routière, l'aviation, la navigation, le dopage, l'économie, la politique, l'aérospatial, la guerre, l'informatique, la pollution, le sport, la télévision et le travail . Nous avons construit un tel ensemble de thèmes dans un objectif de découverte des sujets abordés dans des corpus de textes nous étant peu connus. Les outils MemLabor et ThemeEditor ont été utilisés lors de la construction de ces thèmes. La carte présentée en figure 2 a été construite à partir de ces entrées[10].

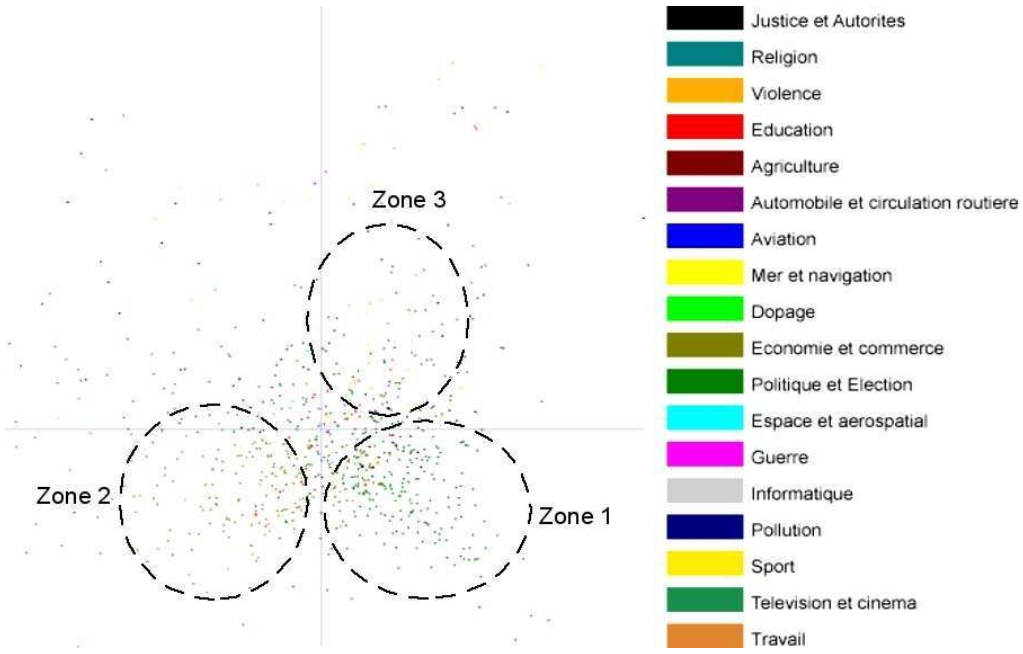

FIG. 2 – Carte thématique du corpus analysé représentant chaque article par un point. Version électronique : `http ://www.info.unicaen.fr/~troy/proxidocs/cartes_classiques/acp1.html`

Chaque point sur la carte représente un article du corpus analysé. La couleur d'un point correspond au thème majoritaire repéré dans l'article représenté[11]. Chaque point est un hyperlien vers l'article représenté. Les zones 1, 2 et 3 ont été marquées manuellement sur la carte afin d'en faciliter l'analyse. La plupart des articles de la zone 1 sont de thème majoritaire Politique et élection

[10] La méthode de comptage relative et la méthode de projection de l'ACP ont été utilisées lors de la construction des cartes présentées dans cet article.

[11] Une légende de couleurs est disponible sur la droite de la carte, l'association d'une couleur à un thème est réalisée par l'utilisateur lors de la construction des thèmes.

alors que la zone 2 contient plus particulièrement des articles de thème majoritaire Économie et commerce. La zone 3, présente en haut et à droite de la carte, contient un petit nombre d'articles de thème majoritaire Guerre.

À partir de la carte précédente, nous avons choisi d'aider l'utilisateur dans son analyse en appliquant une méthode de catégorisation automatique sur les textes de la carte. La carte présentée en figure 3 met en évidence les résultats de cette catégorisation[12]. Chaque groupe de textes est représenté sur la carte par un disque de taille proportionnelle à sa cardinalité. Chaque disque est centré sur le centre de gravité de l'ensemble des points représentant les documents du groupe. La couleur attribuée à ce disque correspond au thème majoritaire repéré dans les textes du groupe. Chaque disque est un hyperlien vers le texte *représentatif* du groupe, c'est-à-dire le texte étant le plus proche de son centre de gravité. Sur cette carte, chaque groupe est caractérisé par les cinq lexies des thèmes les plus fréquentes au sein des textes du groupe.

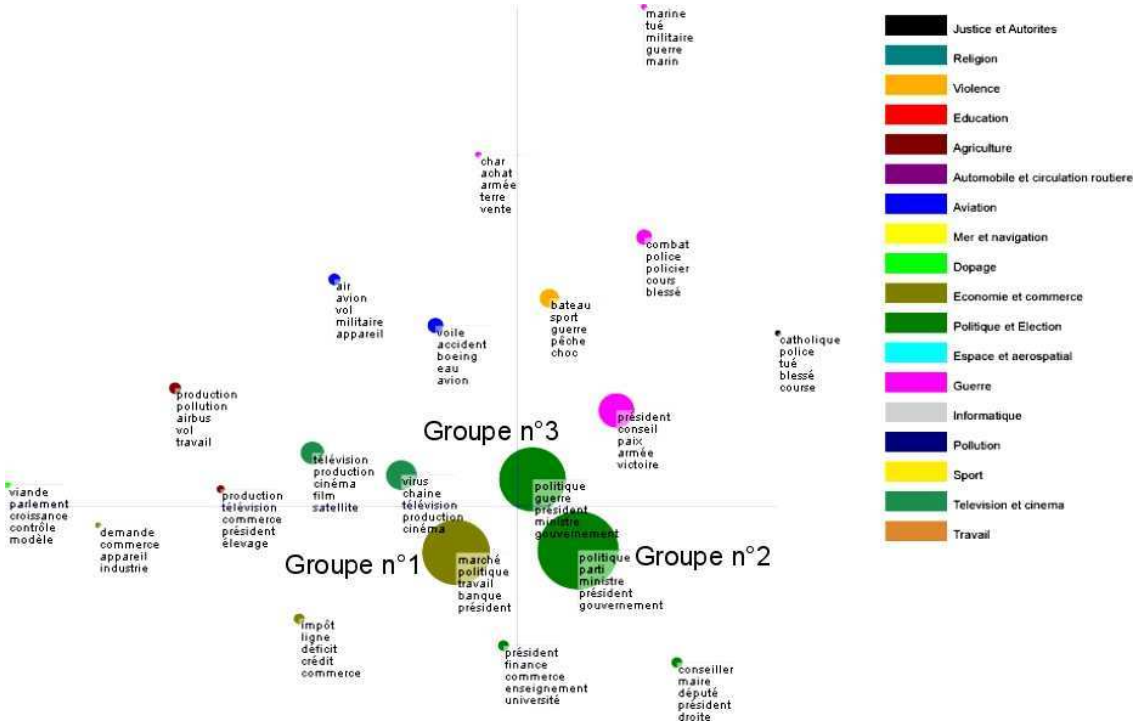

FIG. 3 – Carte thématique mettant en évidence des groupes d'articles. Version électronique : http ://www.info.unicaen.fr/~troy/proxidocs/cartes_categorisation/carte_avec_groupes.htm

La couleur, la taille et la disposition des groupes sur la carte donnent une idée sur les thèmes abordés dans les textes du corpus ainsi que sur leur répartition. En visualisant les textes représentatifs des groupes, l'utilisateur peut avoir une idée plus précise des thèmes abordés dans les textes de chaque groupe. Ainsi, les textes représentatifs des groupes 1 et 2 (groupes marqués manuellement sur la carte, tout comme le groupe 3), traitent respectivement du rachat des parts boursières d'une grande entreprise et des enjeux des futures élections européennes. Il est alors possible de déduire que les thèmes abordés dans ces articles se retrouvent dans les autres textes de leurs groupes respectifs.

Les lexies caractérisant les groupes de textes sur la carte peuvent également aider l'utilisateur dans l'appréhension des thèmes abordés au sein des groupes. Les cartes étant interactives, lorsque l'utilisateur passe avec sa souris sur l'une de ces lexies, celle-ci ainsi que les lexies

[12] Le nombre de groupes empiriquement choisi dans cet exemple est de 20.

identiques caractérisant les autres groupes se colorient en rouge. Cette opération permet ainsi d'observer des lexies communes à plusieurs groupes ou bien des lexies ne caractérisant qu'un seul groupe. Nous pouvons ainsi observer que la lexie politique est commune aux groupes 1, 2 et 3. Par contre, si nous souhaitons différencier ces trois groupes les uns des autres, nous pouvons remarquer que le groupe 3 est le seul à posséder la lexie guerre, le groupe 2 est le seul à posséder la lexie parti et le groupe 1 est le seul à posséder les lexies marché, travail et banque.

Afin d'offrir à l'utilisateur une vision encore plus précise et dynamique de son corpus, nous proposons de tenir compte du temps dans la construction de ses cartes. Pour cela, nous construisons des cartes thématiques à partir du corpus et des thèmes choisis par l'utilisateur sur différentes périodes[13]. Une carte dynamique proposant un enchaînement automatique de ces cartes peut alors mettre en évidence l'évolution des thèmes abordés dans les articles du corpus au fil du temps. A partir du corpus et des thèmes considérés précédemment, une carte dynamique a pu être construite, des extraits de cette carte sur deux périodes sont présentés en figure 4.

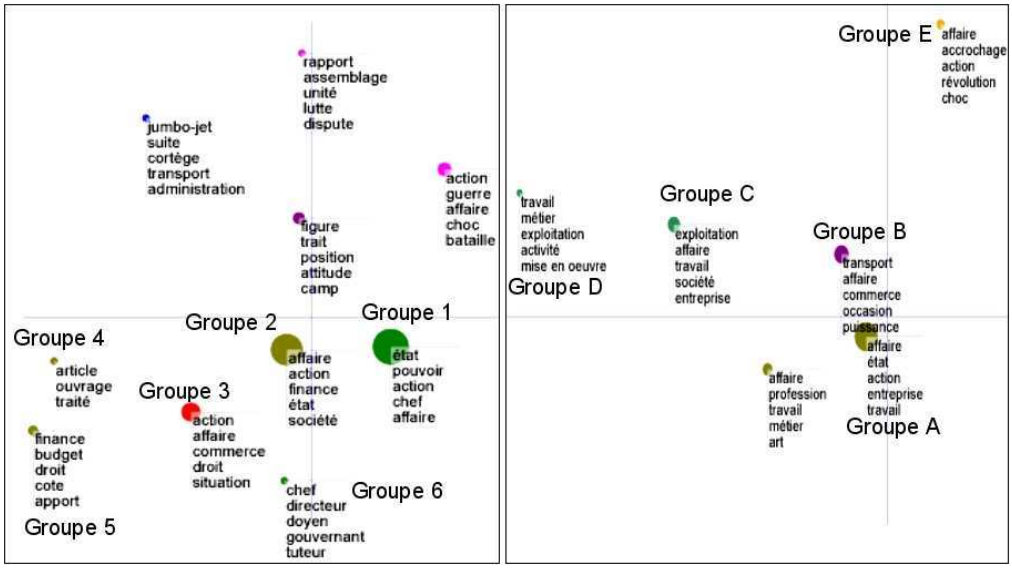

FIG. 4 – Extraits de la carte dynamique globale du corpus. Version électronique : http ://www.info.unicaen.fr/~troy/proxidocs/cartes_temps/carte_dyn_1.html

L'extrait situé sur la partie gauche de la figure met en évidence des groupes d'articles dont la date de publication est comprise entre le 28 janvier et le 27 février 1989. L'extrait de droite représente des groupes d'articles publiés entre le 28 février et le 27 mars. Pour des raisons de lisibilité, la légende de couleurs, identique à celles des cartes des figures 2 et 3, n'est pas rappelée. Pour ces mêmes raisons, certains groupes sont marqués manuellement sur les extraits.

L'extrait de gauche met en évidence deux importants groupes d'articles de thèmes majoritaires Politique et élection et Économie et commerce (groupes 1 et 2, contenant respectivement 28 et 24 articles). Sur la partie en bas et à gauche, un groupe de thème majoritaire Education est également présent (groupe 3). Ce groupe se situe à proximité de groupes de thèmes majoritaires Économie et commerce et Politique et élection (groupes 1, 4, 5 et 6), ce qui laisse penser que les articles contenus dans ce groupe abordent d'une certaine manière ces deux thèmes. Cette idée se confirme en visualisant le texte représentatif du groupe 3, ce dernier abordant des réformes budgétaires du gouvernement sur le système éducatif. La partie de l'extrait située au-dessus de l'axe des abscisses met en évidence des groupes de petite taille (contenant 1 ou 2 articles). En

[13]Dans l'exemple présenté ici, nous nous sommes basés sur la date de publication des articles.

visualisant les articles représentatifs de ces groupes, nous pouvons remarquer qu'ils abordent des sujets d'actualité très ponctuels, tels un crash d'avion et des actes terroristes.

L'extrait de droite met toujours en évidence un important groupe de thème majoritaire Économie et commerce (groupe A, contenant 22 articles). En visualisant la carte dynamique globale du corpus, nous pouvons remarquer qu'un tel groupe est présent tout au long de la période étudiée, ce qui peut laisser penser que le thème de l'économie est constamment abordé dans les articles du corpus analysé. Au-dessus du groupe A se situe un groupe de thème majoritaire Automobile et circulation routière (groupe B). La grande majorité des articles de ce groupe traitent des ventes de voitures en France. Nous pouvons également remarquer la présence de petits groupes (contenant 1 ou 2 articles) de thèmes majoritaires Télévision et cinéma (groupes C et D) et Violence (Groupe E). Les articles représentatifs de ces groupes traitent de sujets d'actualité ponctuels liés à la disparition d'un grand producteur de cinéma (groupes C et D) et au décès d'un jeune boxeur lors d'un combat (groupe E).

La figure 4 commentée précédemment présente deux extraits de la carte dynamique globale retournée à l'utilisateur. Cette carte globale permet entre autres de visualiser les thèmes constamment abordés dans les articles du corpus tout au long de la période étudiée, mais aussi d'observer des thèmes liés à l'actualité abordés de façon plus ponctuels dans les textes.

D'autres possibilités, non détaillées dans cet article, sont également offertes par la plate-forme, telle la possibilité de construire des cartes thématiques en 3 dimensions[14].

4 Conclusions et perspectives

Dans cet article, nous avons présenté les principes de fonctionnement de la plate-forme logicielle ProxiDocs dédiée à la cartographie thématique de corpus de textes. Nous avons illustré les intérêts de cette plateforme sur un exemple précis : la découverte des thèmes abordés dans un corpus d'articles d'un grand quotidien français. Les différentes cartes construites permettent de mettre en évidence les principaux sujets abordés dans les textes de cet ensemble, d'observer des groupes de textes abordant des thèmes proches et de visualiser l'évolution des sujets abordés dans ces articles au fil du temps.

Plusieurs améliorations de la plate-forme ProxiDocs sont actuellement envisagées. D'un point de vue théorique, nous souhaitons intégrer un modèle de représentation des thèmes beaucoup plus fin que celui utilisé jusqu'à présent (dépassant les simples listes de lexies). Le modèle de représentation lexicale envisagé (intitulé LUCIA) est expérimenté depuis plusieurs années au sein de notre équipe (Nicolle et al., 2002; Perlerin, 2004). L'intégration de ce modèle à notre plate-forme permettrait à l'utilisateur de préciser et de structurer les lexies relevant des thématiques de son choix en précisant, pour chacune d'elles, les significations qu'ils jugent pertinentes et appropriées à la tâche qu'il vise. Les cartes ainsi produites devraient révéler des informations plus précises sur le corpus analysé et surtout plus en rapport avec le point de vue de l'utilisateur ou du groupe d'utilisateurs destinataires des cartes.

D'un point de vue applicatif, nous avons commencé le développement de deux composants :
– le métamoteur de recherche ProxiDocs Web interrogeant des moteurs de recherche généralistes (tels Google, Yahoo, etc.) à partir de mots-clés saisis par l'utilisateur et retournant les

[14]Des exemples de cartes en 3 dimensions sont disponibles à l'adresse : http://www.info.unicaen.fr/~troy/proxidocs/cartes_3D

pages proposés par ces moteurs sous la forme de cartes thématiques construites à partir de thèmes choisis et définis par l'utilisateur ;
– et l'outil ProxiDocs Mail réalisant la cartographie dynamique d'un flux de courriers électroniques selon des thèmes choisis et définis par l'utilisateur.

Ces deux outils devraient nous permettre de traiter d'autres types de corpus et ainsi proposer de nouveaux services aux utilisateurs.

Afin de mettre en évidence les intérêts et les limites de notre plate-forme, une expérimentation avec un grand nombre d'utilisateurs nous semble incontournable. Cette expérimentation, et surtout son évaluation, n'est pas sans poser problèmes car il n'est pas question ici de juger (par exemple, en terme de rappel et de précision) si la plate-forme produit des résultats corrects ou non, mais plutôt d'évaluer la façon dont les utilisateurs s'approprient l'outil chacun à leur façon selon leurs buts. Ce n'est donc pas simplement le logiciel qu'il faut évaluer mais le couple *outil-utilisateur*. À travers une telle expérimentation, de nouveaux besoins devraient émerger, ceci nous persuadant à toujours mieux instrumentaliser la dimension intertextuelle de la sémantique des langues.

Références

Benzécri J.P. (1980), *L'analyse des données - tome 2 : l'analyse des correspondances*, Éditions Bordas.

Beust P. (2002), Un outil de coloriage de corpus pour la représentation de thèmes, Actes des *6èmes Journées internationales de l'Analyse statistique de Données Textuelles*.

Bouroche J.M. et Saporta G. (1980) *L'analyse des données*, Collection Que sais-je ?, PUF.

Chung W., Chen H. et Numaker J.F.Jr. (2002) Business Intelligence Explorer : A Knowledge Map Framework for Discovering Business Intelligence on the Web, Actes de la *36ème HICSS*.

Lelu A. et Aubin S. (2001) Vers un environnement complet de synthèse statistique de contenus textuels, Présentation au séminaire *Association pour la mesure des sciences et des techniques* du 13/11/2001.

Mokrane A., Arezki R., Dray G. et Poncelet P. (2004) Cartographie automatique du contenu d'un corpus de documents textuels, Actes des *7èmes Journées internationales de l'Analyse statistique de Données Textuelles*, pages 816-823.

Nicolle A. (1996), L'expérimentation et l'intelligence artifi cielle, *Intellectica*, numéro 22, pages 9-19, Association pour la Recherche Cognitive (ARC).

Nicolle A., Beust P. et Perlerin V. (2002), Un analogue de la mémoire pour un agent logiciel interactif, *In Cognito*, numéro 21, pages 37-66.

Perlerin V. (2002), MemLabor, un environnement de création, de gestion et de manipulation de corpus de textes, Actes de *TALN / RECITAL 2002*, pages 507 à 516.

Perlerin V. (2004), *Sémantique légère pour le document : assistance personnalisée pour l'accès au document et l'exploration de son contenu*, Thèse d'informatique de l'Université de Caen.

Pichon R. et Sébillot P. (1999), Différencier les sens des mots à l'aide du thème et du contexte de leurs occurrences : une expérience, Actes de *TALN 1999*, pages 279-288.

Roy T. et Beust P. (2004), ProxiDocs, un outil de cartographie et de catégorisation thématique de corpus, Actes des *7èmes Journées internationales de l'Analyse statistique de Données Textuelles*, pages 978-987.

Spinat E. (2002), Pourquoi intégrer des outils de cartographie au sein des systèmes d'information de l'entreprise ?, Colloque *Cartographie de l'Information*, Paris.

Thlivitis T. (1998), *Sémantique interprétative intertextuelle : assistance anthropocentrée à la compréhension des textes*, Thèse d'informatique de l'Université de Rennes I.

W3C (2001), *Scalable Vector Graphics (SVG)*, `http://www.w3.org/TR/SVG/`.