Mathématiques et sciences humaines
Mathematics and social sciences

197 | Printemps 2012
Catégories, classification, complexité, consensus…
Autour des travaux de Jean-Pierre Barthélemy

Analyse textuelle de travaux de Jean-Pierre Barthélemy

Textuel analysis of a part of Jean-Pierre Barthélmy's works

Marc Le Pouliquen

Édition électronique
URL : http://journals.openedition.org/msh/12170
DOI : 10.4000/msh.12170
ISSN : 1950-6821

Éditeur
Centre d'analyse et de mathématique sociales de l'EHESS

Édition imprimée
Date de publication : 22 avril 2012
Pagination : 33-45
ISSN : 0987-6936

Référence électronique
Marc Le Pouliquen, « Analyse textuelle de travaux de Jean-Pierre Barthélemy », *Mathématiques et sciences humaines* [En ligne], 197 | Printemps 2012, mis en ligne le 02 mai 2012, consulté le 05 mai 2019. URL : http://journals.openedition.org/msh/12170 ; DOI : 10.4000/msh.12170

Math. & Sci. hum. / Mathematics and Social Sciences (50e année, n° 197, *2012*(1), p. 33–45)

ANALYSE TEXTUELLE DE TRAVAUX DE JEAN-PIERRE BARTHÉLEMY

Marc LE POULIQUEN[1,2]

Jean-Pierre Barthélemy nous a quitté le 21 juin 2010. À l'initiative d'Olivier Hudry, une journée d'hommage s'est tenue le 21 juin 2011. Ce fut l'occasion de se souvenir et de redécouvrir la large bibliographie de Jean-Pierre Barthélemy. Cet article est dédié à cet homme de culture, de sciences et de lettres.

RÉSUMÉ – *Dans cet article, nous allons utiliser les représentations factorielles et arborées pour visualiser une partie des travaux de Jean-Pierre Barthélemy.*

Chercheur pendant quarante ans, Jean-Pierre Barthélemy aimait illustrer ses livres et ses articles par des classifications textuelles effectuées sur le vocabulaire des œuvres de Giraudoux ou celles plus polémiques de Molière et Corneille. Il était toujours étonné des résultats obtenus par les classifications automatiques dans le domaine textuel.

Pour lui rendre hommage, nous avons réalisé plusieurs classifications d'une soixantaine de ses articles scientifiques en français ou en anglais, en fonction des thèmes de recherche. Pour réaliser ces classifications, nous avons utilisé l'outil BI-Qnomis, un logiciel pour l'analyse textuelle factorielle ainsi que les méthodes arborées que Jean-Pierre Barthélemy a contribué à populariser dans les années 1980.

MOTS CLÉS – Analyse arborée, Analyse des correspondances, Analyse de données textuelles, Bibliographie, Métaclé

SUMMARY– Textual analysis of a part of Jean-Pierre Barthélemy's works

In this article, we will study a part of Jean-Pierre Barthélemy's works using textual analysis methods. Formerly a researcher for almost forty years, he used textual classifications of Giraudoux's and Zola's texts as examples to illustrate both his papers and his books. He was often surprised by the results he obtained.

As a tribute to Barthélemy, we have carried out a classification of about sixty of his scientific papers, both those written in French and those written in English. To carry out this classification, we used BI-Qnomis which is a factorial textual analysis software and algorithm for tree reconstruction which Jean-Pierre Barthélemy contributed to make popular in the 1980s.

KEYWORDS – Bibliography, Correspondence analysis, Metakey, Textual data analysis, Tree analysis.

[1]Laboratoire en Sciences et Technologies de l'Information de la Communication et de la Connaissance (LabSTICC), UMR CNRS 31922, Telecom Bretagne, Technopôle Brest-Iroise, CS83818, 29238 Brest Cedex 3, marc.lepouliquen@enst-bretagne.fr

[2]IUP Génie Mécanique et Productique, Université de Bretagne Occidentale, 6 avenue Le Gorgeu, CS93837, 29238 Brest Cedex 3

1. INTRODUCTION

Dans cet article, nous allons examiner, par des méthodes de l'analyse textuelle, une partie de la bibliographie de Jean-Pierre Barthélemy.

Professeur à Telecom Bretagne pendant presque vingt ans, il se plaisait à agrémenter son cours de classification, ainsi que les exemples de ses livres (cf. [Brucker, Barthélemy, 2007]) et de ses articles (cf. [Barthélemy, Luong, 1998]) par des classifications sur les œuvres de Giraudoux ou de Zola. On se souvient aussi de la polémique au sujet de l'attribution des œuvres de Molière à Corneille pour laquelle Jean-Pierre s'insurgeait dans un article du monde du 10/06/2003 :

> *On ne peut pas faire passer pour des statistiques inférentielles, avec lesquelles on peut éprouver des hypothèses, des statistiques descriptives, d'abord destinées à faire réfléchir des spécialistes.*

Il était cependant toujours étonné des résultats obtenus par les classifications automatiques dans le domaine textuel.

Si la classification et l'analyse textuelle faisaient partie de son domaine de recherche, il œuvrait sur de nombreux autres fronts comme les mathématiques discrètes avec les treillis de Galois, les sciences cognitives, l'aide à la décision, les mathématiques psychologiques... C'est donc un ensemble de travaux scientifiques d'envergure et de diversité que l'on retrouve à travers sa bibliographie.

Nous allons essayer, à travers deux outils, de visualiser et de retrouver une classification, en thème de recherche, d'une partie des articles scientifiques auxquels il a contribué pendant trente-cinq ans. La première étude consiste à utiliser l'outil BI-Qnomis[3], un logiciel pour l'analyse textuelle développé par Michel Kerbaol (cf. [Kerbaol *et al.*, 2006]). Une fois les articles convenablement formatés, le logiciel utilise l'Analyse Factorielle des Correspondances (AFC) introduite en France par Benzécri [1973]. La première étape consiste à constituer un tableau de données avec, en ligne, les articles étudiés et, en colonne, les mots les plus fréquents et les plus cooccurrents. Cette étude nous permet d'abord d'avoir un premier aperçu des thèmes qui ressortent sur les premiers axes factoriels.

À partir du tableau obtenu précédemment, nous pouvons alors réaliser une matrice de dissimilarités, voire de distances, et utiliser les représentations arborées si chères à Jean-Pierre. Pour cette seconde classification, nous pourrons utiliser *l'algorithme des groupements* développé avec Luong [1988], une méthode de score dérivée du célèbre *AddTree* de Sattah et Tversky [1977].

2. LE CORPUS

Pour la constitution du corpus, il n'y a pas de thème précis mais un auteur, Jean-Pierre Barthélemy. Les articles à récupérer ne font donc pas partie d'un domaine précis mais concernent plusieurs disciplines, ce qui complique la recherche bibliographique. Partant d'un certain nombre d'articles précédemment obtenus, de ceux

[3]BI-Qnomis est un logiciel conçu par Michel Kerbaol et Joël Josse, Log INSERM © 1979 – 1987 – 1993.

récupérés sur les bases en ligne et de ceux que nous ont envoyés les co-auteurs, il a été possible d'établir un premier corpus. Par la consultation des bibliographies de ce premier corpus et de celle des nouveaux articles, une liste de 80 articles écrits entre 1974 et 2009 a été réalisée. Une liste plus complète est proposée à la fin de ce numéro spécial et collectée par B. Monjardet.

Sur les 80 articles listés, nous disposons d'une version numérique pour 51 d'entre eux. Pour les autres, nous ne détenons aucune version ou une version papier dont la reconnaissance optique demandent malheureusement une révision complète de l'article qui n'a pas été réalisée.

En analyse des données textuelles, la mise en forme et le nettoyage des fichiers sources sont une étape fastidieuse mais nécessaire. L'utilisation de BI-Qnomis impose un pré-traitement des documents textes d'origines diverses et dont les caractères accentués ou spéciaux ainsi que les formules mathématiques sont à réviser. De plus, il est nécessaire de découper les articles en paragraphes car nous constatons généralement que les coocurences de mots au sein d'un paragraphe sont porteuses de sens, tandis que celles au sein d'un article complet le sont beaucoup moins. Une procédure en Perl permet un nettoyage correct et une mise en forme adéquate des fichiers.

Finalement, nous disposons de deux corpus, un premier en anglais composé de 30 articles et le second en français avec 21 articles que nous allons pouvoir analyser. Les articles sont nommés en utilisant le nom du second auteur (s'il y en a un) ou Barthélemy (s'il est le seul auteur) et complétés de l'année ainsi que de bis (s'il y en a plusieurs). Par exemple, l'article qui suit est nommé Gusho09 voire Gush09 pour des raisons d'affichage :

Jean-Pierre Barthélemy & Gentian Gusho (2009), "On the stability of hierarchical classification: Qualitative approaches", *Mathematical and Computer Modelling* 50(3-4), p. 329-332.

3. UTILISATION DES MÉTHODES FACTORIELLES

En France, l'analyse factorielle des correspondances (AFC) est une méthode très classique pour décrire des tables de contingence. L'AFC a été développée, il y a presque 50 ans, par J.-P. Benzecri et B. Escofier Cordier ; dans un contexte linguistique, les premières études ont porté sur les tragédies de Racine. C'est une méthode algébrique qui utilise la décomposition dans la base des vecteurs propres d'une matrice pour réduire la dimension du problème et permettre une analyse plus facile. Le résultat est une représentation simultanée des lignes (ici, les articles) et des colonnes (ici, des mots) dans un plan dont on peut déterminer la qualité par les contributions des mots ou des articles à l'inertie d'un axe.

3.1. Lexique et métaclés

Le premier problème est d'obtenir une liste intéressante de mots que nous appelons lexique permettant de réaliser la matrice de contingence.

Un premier lexique est constitué grâce au logiciel BI qui référence l'intégralité des mots figurant dans tous les articles. Après suppression des mots-outils, le calcul de la fréquence et de la cooccurrence de chacun des mots du lexique est établi successivement pour l'ensemble des articles. Un tri par ordre décroissant permet de constituer un fichier des premiers mots pour établir le tableau croisé.

M. Kerbaol appelle métaclé les groupes de mots dont les contributions sont très élevées sur un axe. Nous avons deux métaclés par axe, une positive et une négative, qui seront employées pour l'exploration du corpus. La métaclé peut se définir comme une association de mots qui détermine un axe sur lequel on peut retrouver les articles qui lui sont liés. Un mot peut se trouver dans plusieurs métaclés et une métaclé peut définir un mélange de thèmes. Il faut donc identifier les axes les plus pertinents (associés à un thème) par les métaclés, opération manuelle qui consiste à observer les différents tableaux de métaclés pour identifier les plus intéressants. L'AFC permet de visualiser graphiquement ces métaclés sous la forme d'une suite de plans et le thème dégagé par l'association de mots provoquant une classification naturelle des articles. Une fois la classification des mots et des articles réalisée, l'analyse du corpus peut commencer en identifiant et en nommant les regroupements pertinents.

3.2. Application au corpus anglais

Dans le cas du corpus anglais, nous obtenons 4693 mots différents dans le corpus. Nous avons isolé les 666 mots les plus fréquents et cooccurrents, que nous avons répartis en 60 classes de 20 mots, les métaclés. L'étude peut alors commencer.

La première présentation permet de trier les axes intéressants, c'est-à-dire ceux pour lesquels les métaclés semblent représenter un thème. Sur la Figure 1, qui correspond à la métaclé négative liée au premier axe factoriel, nous obtenons un groupe de mots liés avec les sciences cognitives (*strategies*, *cognitive*, *maker*, *process*, *expert*, *control*) et avec l'aide à la décision (*decision*, *expert*, *Lenca*, *rule*). C'est toute une partie de la recherche à laquelle Jean-Pierre Barthélemy participait à partir de 1995 avec un certain nombre de collaborateurs comme Lenca, Coppin, Le Saux, Kala Kamdjoug, Legrain, Bisdorff...

La liste des métaclés est aussi accessible par l'interface et permet de visualiser les articles dans lesquels figurent les mots de la métaclé. Ainsi, la Figure 2 présente un des paragraphes d'un article de Kala Kamdjoug *et al.* [2007] et surligne les mots concernés.

L'analyse de correspondance de la table de contingence fournit une autre visualisation des sujets. Sur le premier plan factoriel visible sur la Figure 3, l'interface Qnomis permet l'affichage simultané des projections (les mots et les articles) ; ici, nous n'avons affiché que les mots pour ne pas surcharger le graphique. Nous reconnaissons à gauche la métaclé n° 1. À droite, en haut, nous retrouvons les travaux de Bécassine[4] en classification à partir de 2000 avec Brucker, Osswald, Gusho, ... Il s'agit de l'analyse et de l'élaboration de modèles de classification (en particulier

[4]Travaux qui ont débouché sur la publication du livre *Éléments de classification* de Brucker et Barthélemy introduit par un savoureux extrait de L'Enfance de Bécassine (cf. [Brucker, Barthélemy, 2007]).

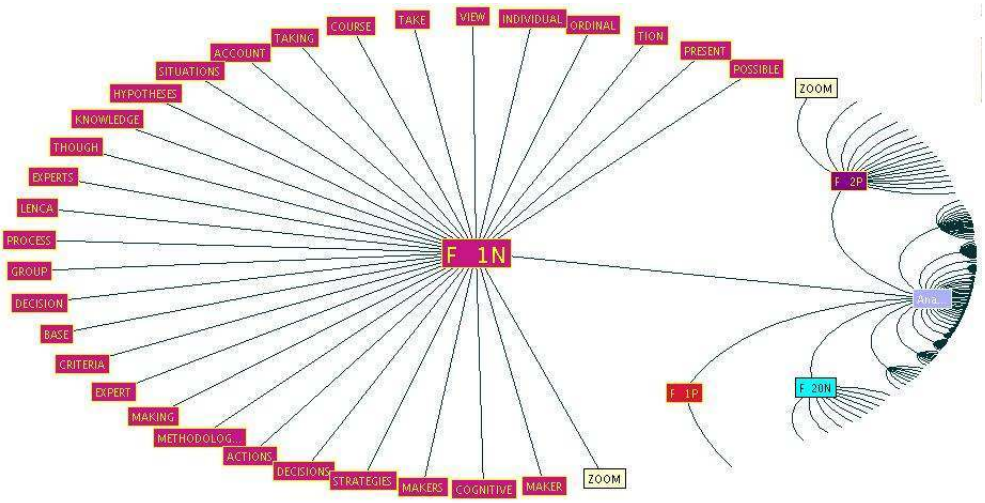

FIGURE 1. Mots partitionnés en arbre, représentation de la métaclé 1 négative

Liste des textes associes a la METACLE 1 Negative X 1 edition lineaire

" " ACCOUNT, ALREADY, approach, COMMON, CONTEXT, DECISIONS, ELEMENTARY, GROUP, HYPOTHESES, LESS, MADE, MAIN, MAKER, MUST, POINTS, POSSIBLE, PROCESS, TAKE, THOUGH, VIEW,

TIALIGN - ID028 Kala07

ALINEA - that same context less preferable than that already beenrejected decisions must made group the decision-makers involved cooperate bear mind that they working towards success common goal even though they have conicting points view objective this st imitate closely possible expert deci-sion-maker behaviourfirst approach should take into account cognitive constraints amongstthes cognitive hypotheses the process decision-making decision making viewed articulation elementary strateg

IDEN - 028

AUTEUR - KALA

TITRE - Assessment of actions in a multi-actor and multicriteria framework: application to the refunding of microfinance institutions

VIE - 2007

SAISI - Comput Econ (2007)

FIGURE 2. Métaclé et article associé

emboîtés) à partir de théorèmes de bijection entre les systèmes de classes et des modèles de dissimilarités. À droite, en bas, nous retrouvons les travaux sur les treillis de Galois et les médianes réalisés avec Monjardet et Leclerc dans les années 80.

Si l'on s'intéresse au plan selon les axes 3 et 4 sur la Figure 4, on constate qu'il représente les articles des années 90. On retrouve ainsi :

– à droite, les travaux réalisés avec Bandelt et Constantin sur les graphes médians et sur le consensus ;

– en haut à gauche, les travaux sur les heuristiques avec les termes *tabu*, *optimal*, *heuristic*, *problem*, *cost* réalisé en collaboration avec De Amorin, Mullet, Hudry et Guénoche ;

– en bas, les mots *ultrametrics*, *dissimilarity*, *weak* font référence à la classification du groupe Bécassine (cf. précédemment).

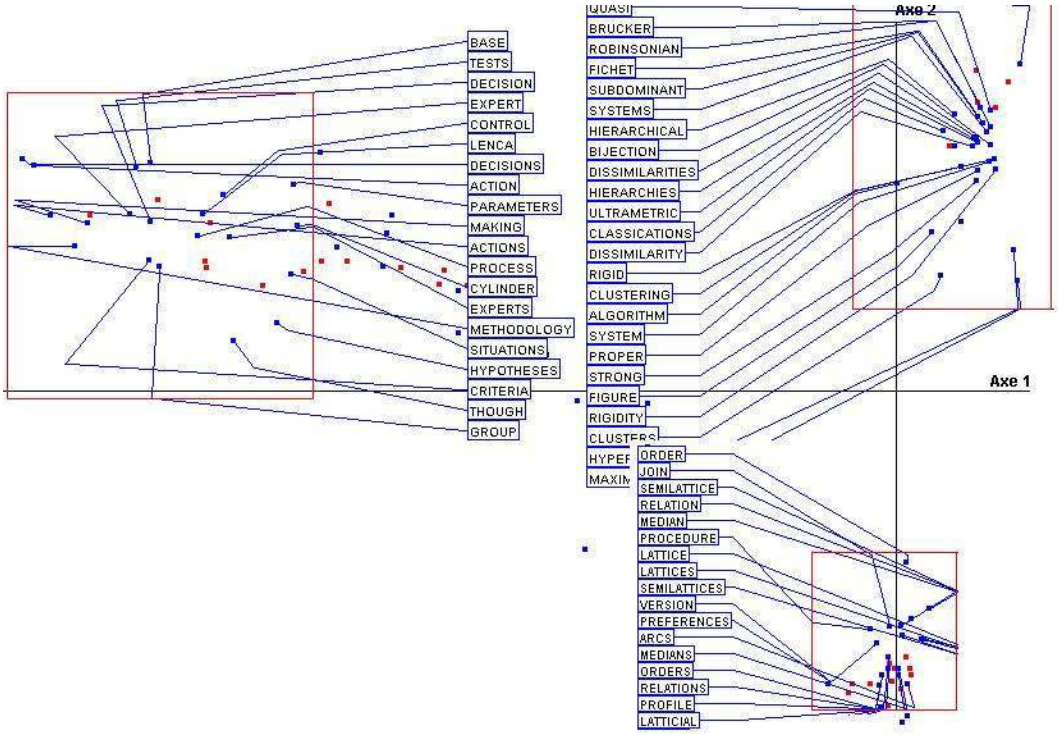

FIGURE 3. Plan factoriel selon les axes 1 et 2

On peux noter que les mots au centre du plan *concensus* et *median*[5] sont des mots polysémiques; par exemple le terme *consensus* est utilisé en décision (*consensus function*) et en classification (*consensus tree*).

3.3. Application au corpus français

Dans le cas du corpus français, avec 20 articles, nous obtenons 4477 mots différents sur le corpus. Nous avons isolé les 494 mots qui constituent les métaclés. On peut par exemple découvrir sur la métaclé 1 positive (cf. Figure 5), le travail d'analyse textuelle sur les arbres de la filiation appelés *stemma codicum* sur un corpus de manuscrits sanskrits, travail réalisé avec Le Pouliquen.

L'affichage simultané des projections sur les axes 1 et 2 est visible sur la Figure 6. Sur l'axe 1, nous y reconnaissons à nouveau les travaux sur l'aide à la décision visibles à gauche et les articles consacrés à l'analyse textuelle sur la filiation de manuscrits à droite. En haut, les termes sont liés aux travaux de Jean-Pierre dans le domaine des graphes médians, dont la plupart des articles datant des années 80 sont écrits seul ou avec Monjardet. Ils sont complétés par un article avec Hudry en 2006. En bas, on retrouve les articles de Jean-Pierre sur les propriétés métriques réalisés entre 1975 et 1980 c'est-à-dire peu après sa première thèse obtenue en 1971 dans le domaine des catégories et de la logique sous le titre de *Esquisses pointées* et sous la direction de Charles Ehresmann.

Le plan selon les axes 3 et 4 de la Figure 7 est lui aussi intéressant. On visualise

[5]Lors de la journée d'hommage, B. Fichet me faisait remarquer qu'il était normal que *median* soit central !

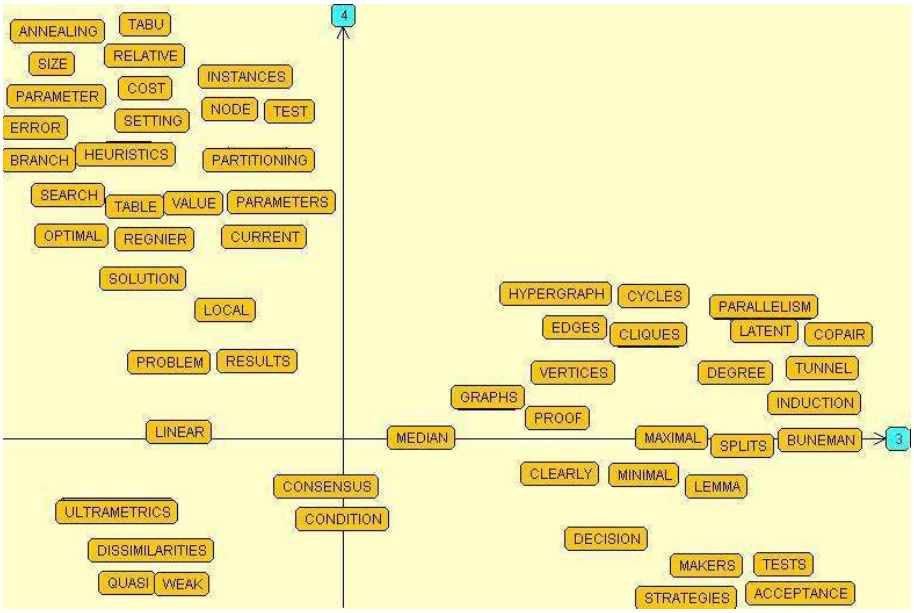

FIGURE 4. Plan factoriel selon les axes 3 et 4

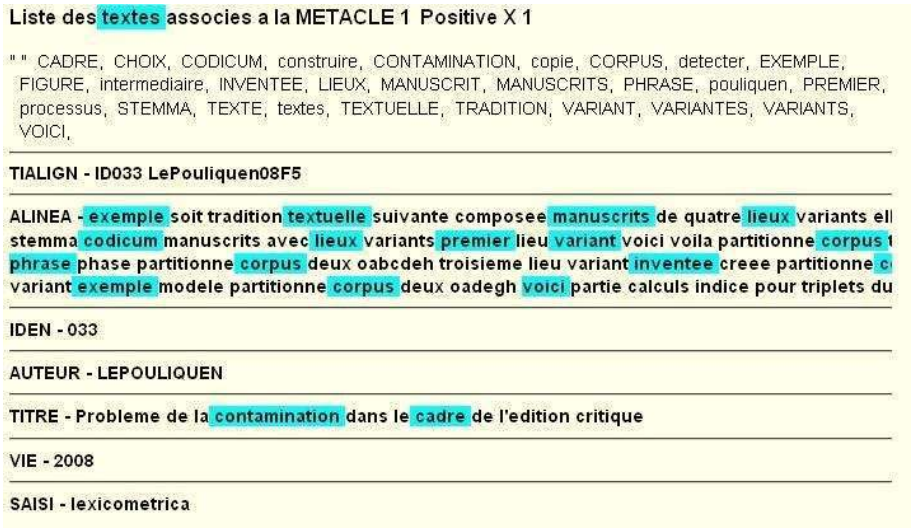

FIGURE 5. Métaclé 1 positive : analyse textuelle

toujours l'aide à la décision, la filiation de manuscrits et le domaine des graphes médians dans la partie gauche de la projection. Dans la partie droite, on peut aussi y voir les travaux de Jean-Pierre Barthélemy sur les sciences cognitives avec De Glas, Desclé et Petitot à travers les termes *morphodynamique*, *système*, *sémantique* et *langage*. Plus au centre, à travers les mots *voisinage* et *topologie*, c'est la collaboration avec le laboratoire BCL[6] (Bases, Corpus, Langage) sur la définition de la topologie textuelle établie avec S. Mellet et D. Longrée.

[6]Jean-Pierre Barthélemy a noué des liens privilégiés avec le Laboratoire BCL de Nice et plus particulièrement avec X. Luong, un de ses anciens thésards, et E. Brunet, le père d'HYPERBASE, un logiciel d'analyse textuelle.

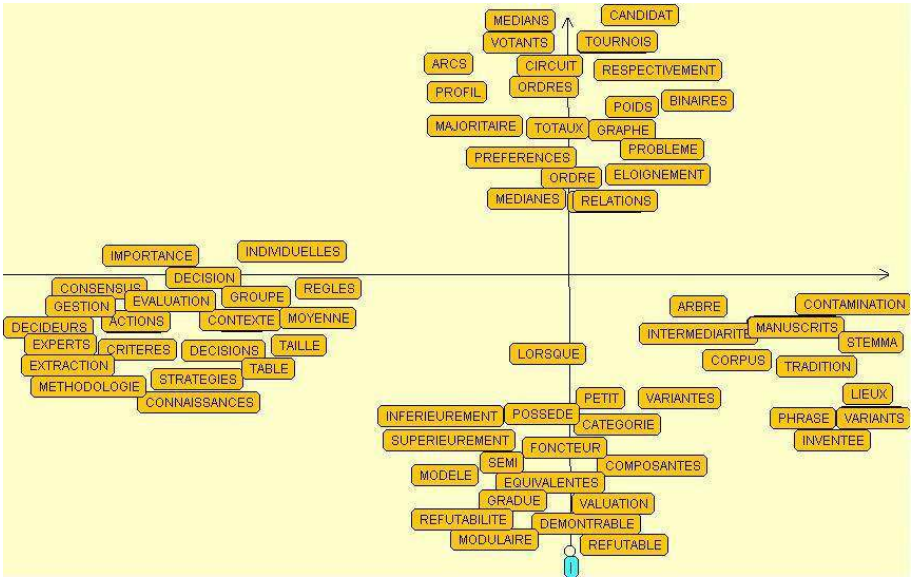

FIGURE 6. Plan factoriel selon les axes 1 et 2

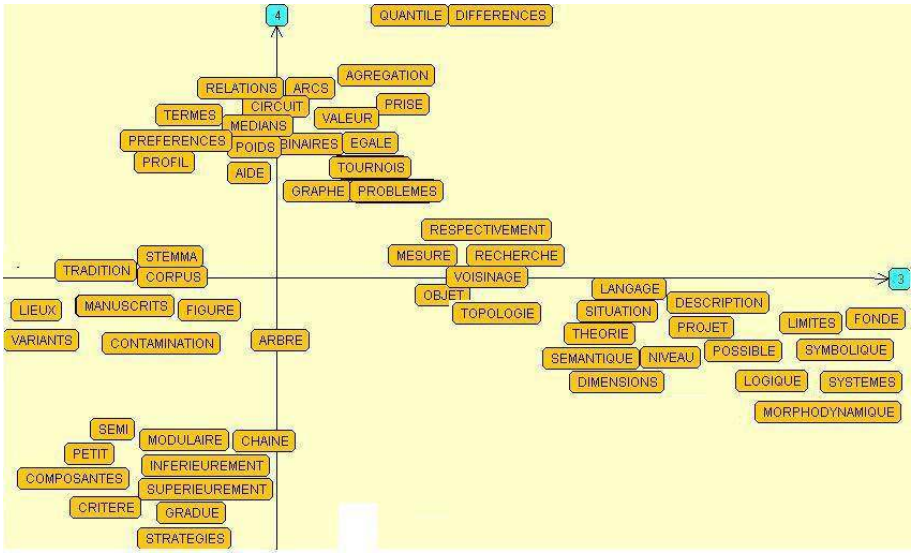

FIGURE 7. Plan factoriel selon les axes 3 et 4

4. UTILISATION DES MÉTHODES ARBORÈES

L'usage de méthodes mathématiques pour traiter de données textuelles a une longue tradition. On se souvient que c'est en étudiant Pouchkine que Markov a élaboré ses chaînes.[...] *Les distances textuelles permettent d'évaluer et de représenter la proximité ou l'éloignement entre deux ou plusieurs textes en fonction de divers paramètres (dont le plus familier et le plus fréquemment étudié est leur lexique). Les mathématiques sont ici beaucoup plus simples ; elles font appel à quelques rudiments sur les espaces métriques discretssouvent couplés avec des méthodes de représentation : calcul statistique, analyses factorielles, multidimensional scaling et représentation arborée.*

C'est un extrait du prolégomène de l'article Prenons nos distances pour comparer des textes, les analyser et les représenter écrit avec X. Luong et S. Mellet une quinzaine d'années après avoir participé activement au développement des méthodes arborées. On remarque tout au long de l'article le plaisir évident et l'expérience, voire l'expertise, que Jean-Pierre Barthélemy et ses co-auteurs ont acquises dans le domaine de la classification textuelle. Il était donc légitime d'utiliser l'algorithme emblématique développé avec X. Luong dans les années 80 pour obtenir une classification arborée des articles collectés.

Dans cette partie, nous allons en premier lieu construire une distance (voire une dissimilarité) entre les différents articles dont nous disposons. À partir de cette matrice de distances, nous allons utiliser *l'algorithme des groupements* développé avec Luong [1988] pour obtenir des représentations arborées.

4.1. Construction de la distance

À partir du tableau booléen de présence/absence avec en ligne, les articles étudiés et en colonne, les mots du lexique obtenu précédemment, nous pouvons utiliser la distance de Jaccard [1912] pour obtenir une table de distance entre les articles (cf. Table 1).

	Amorim92	Bandelt84	Barthelemy82	Barthelemy88	...
Amorim92	0	0,867	0,814	0,755	...
Bandelt84	0,867	0	0,855	0,839	...
Barthelemy82	0,814	0,855	0	0,824	...
Barthelemy88	0,755	0,839	0,824	0	...
⋮	⋮	⋮	⋮	⋮	

Table 1. Extrait de table de distance de Jaccard entre articles

4.2. Méthodes des groupements de Luong

La représentation arborée d'une matrice de distance consiste à déterminer un arbre dont les arêtes entre les sommets sont telles que leur longueur soit au plus proche de la distance d'origine. La méthode des groupements de Luong s'inscrit dans le cadre des heuristiques liées à la représentation arborée au même titre que *ADDTREE* de Sattah et Tversky [1977] et NJ (Neighbours Joining) de Saitou et Nei [1987] pour les plus célèbres.

On considère ici que chaque article correspond à un sommet de l'arbre à construire. Les sommets sont de deux types, les feuilles de l'arbre (c'est-à-dire les sommets reliés à l'arbre par une seule arête) ou les nœuds (c'est-à-dire les sommets internes de l'arbre). La méthode part d'une matrice de distance D entre les différents articles. C'est une méthode itérative qui construit des groupements de sommets (au départ les articles) autour d'un nœud qui correspond soit à l'un des articles existants, soit à un nœud de construction. Pour chaque groupement, on ajoute des arêtes entre les sommets du groupement et le nœud précédemment construit, puis on supprime de D les sommets du groupement en ne conservant ou en ne rajoutant

que le nœud de construction. Les distances entre les sommets restants sont alors calculées et on effectue une nouvelle itération tant qu'il reste au moins trois sommets. On réalise alors un traitement spécifique pour les derniers sommets. Si l'on utilise la table réduite 1, on obtient alors l'arbre de la Figure 8.

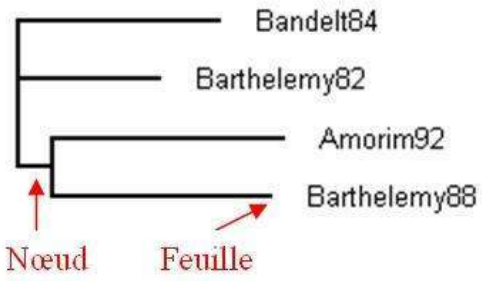

FIGURE 8. Exemple d'arbre inféré avec les données extraites de la table 1

4.3. Application au corpus anglais

L'arbre obtenu sur la Figure 9 ne semble pas pertinent par rapport aux groupements d'articles déjà obtenus par les méthodes factorielles. On ne retrouve pas du tout les classifications explicites en terme de domaine de recherche. Si l'on observe la matrice de distance, les valeurs se regroupent toutes avec une moyenne de $0,85$ et un écart type de $0,074$ comme on peut l'observer sur la Figure 10. Les distances sont trop proches pour être pertinentes dans la construction de l'arbre ; le nombre important de zéros dans la matrice booléenne de présence/absence engendre ce problème.

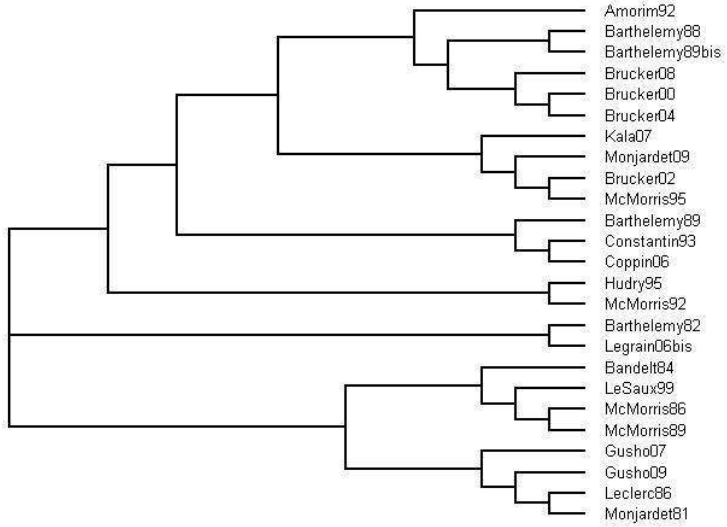

FIGURE 9. Arbre obtenu par les groupements de Luong et visualisé avec TreeView de Page [1996]

Remarques.

1. La distance qui est construite entre les deux masques binaires n'est pas une distance entre les deux articles puisque le masque binaire est construit par un codage avec perte par l'intermédiaire du lexique. La comparaison globale de nombreux articles de thèmes différents à travers les mots du lexique semble être inadéquate.

2. Les différents essais avec d'autres distances comme celle de Hamming (1950) ou avec des indices de similarités comme celui de Dice ou de Russel-Rao présentés chez Lerman [1970] ne donnent pas de résultats plus convaincants.

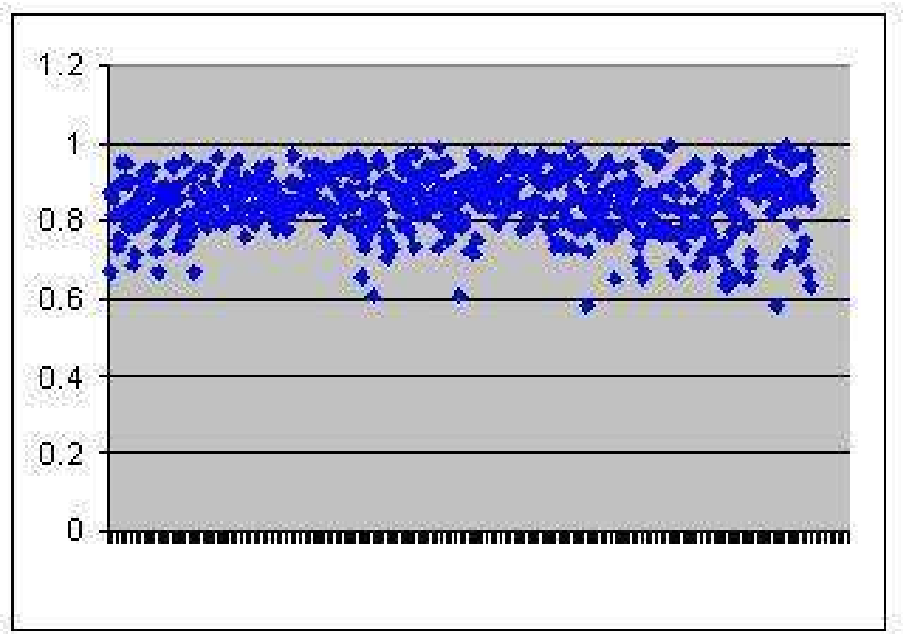

FIGURE 10. Représentation des 600 distances obtenues par le calcul de Jaccard

4.4. Application au corpus français

Dans le cas du corpus français, l'arbre est bien plus pertinent (cf. Figure 11). En prenant moins d'articles, des articles plus homogènes en taille et un lexique plus restreint, on visualise bien les regroupements des articles sur la filiation de manuscrits avec M. Le Pouliquen, des travaux avec le Laboratoire BCL de Nice avec X. Luong et S. Mellet ainsi que ceux dans le domaine des graphes médians dans la partie gauche de l'arbre.

Dans ce cas, l'analyse arborée permet d'obtenir une représentation graphique d'une vingtaine d'articles. Nous n'avons pas réussi à construire une distance, voire un indice de dissimilarité qui permet la visualisation complète et pertinente d'un grand nombre d'articles.

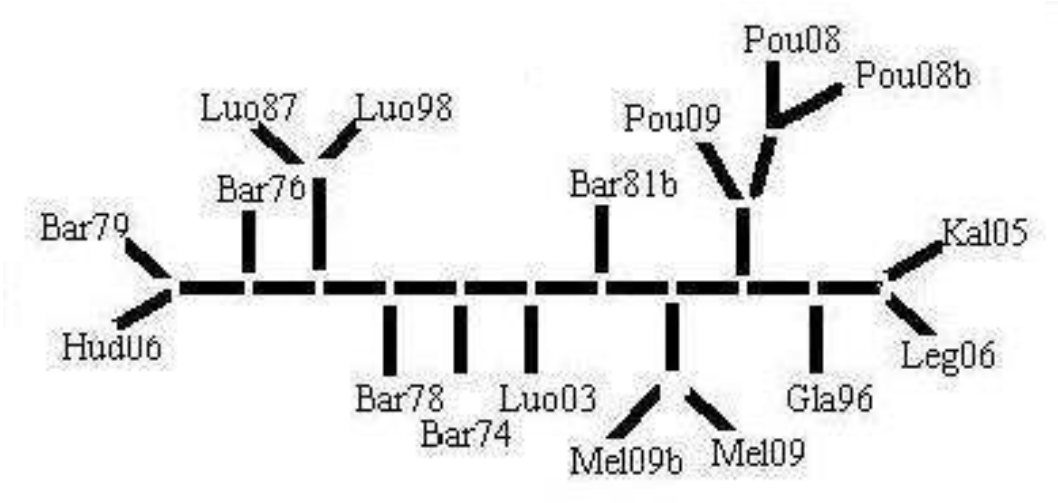

FIGURE 11. Arbre obtenu par les groupements de Luong sur le corpus français

5. CONCLUSION

Les résultats de nos analyses textuelles informatisées permettent à l'utilisateur d'obtenir rapidement des classifications en thèmes de recherche, ceux-ci étant représentés par des listes de mots-clé, les métaclés. Les représentations plus globales sous forme d'arbres ne semblent pas aussi convainquantes que les plans factoriels, les distances utilisées ne discriminent pas correctement l'ensemble des articles.

Même si l'accessibilité des articles scientifiques a fait de gros progrès avec l'arrivée d'Internet, les outils de bibliographie doivent évoluer vers des représentations graphiques adaptées (plan factoriel, arborée, treillis, graphes, ...) facilitant la réalisation de l'état de l'art d'un domaine de recherche, la production d'une bibliographie personnelle ou l'élaboration d'une veille scientifique. Cependant, de nombreuses difficultés demeurent comme la possibilité de travailler sur un corpus multilingue, les problèmes d'harmonisation des données (par exemple : les codages sont divers et variés même au sein des documents pdf) ou la création de représentations plus globales (visualisation simultanée d'un grand nombre d'articles).

Plus personnellement, ce travail m'a permis de voir la diversité et la portée des travaux scientifiques réalisés par Jean-Pierre Barthélemy pendant quarante années. Il est bien sûr impossible de résumer par quelques mots cet apport, mais si l'on voulait faire une tentative, on pourrait essayer avec un *Tag-cloud*. Un *Tag-cloud* ou *nuage de mots-clé* est une visualisation planaire des mots-clé où la taille des fontes est proportionnelle à la fréquence des mots dans les articles. Celui de la Figure 12, réalisé sur les principaux mots des deux lexiques français et anglais, semble convenable.

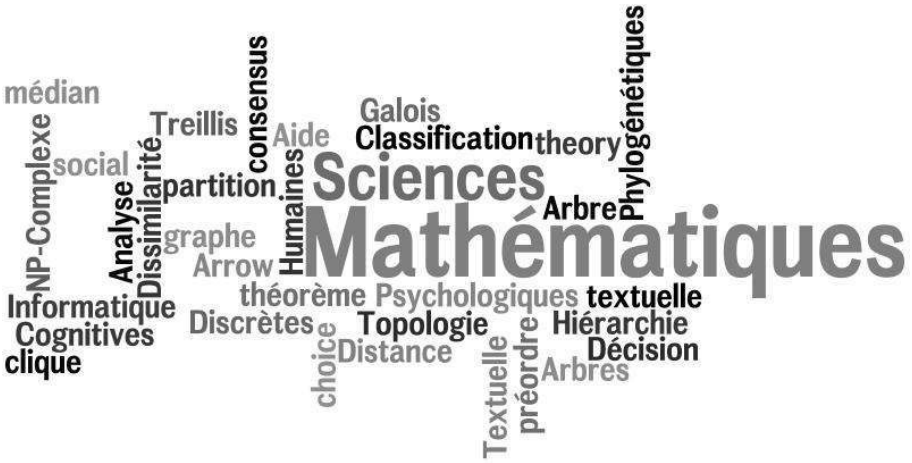

FIGURE 12. *Tag-Cloud* des principaux mots des lexiques francisés

BIBLIOGRAPHY

BARTHÉLEMY J.-P., LUONG X. (1998), « Représenter les données textuelles par des arbres », *JADT98, actes des* 4e *Journées internationales d'analyse de données textuelles*, Université de Nice, UMR 6039, p. 49-70.
http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt1998/barth.htm

BENZÉCRI J.-P. (1973), *Analyse des données*, vol. 1 : *La taxinomie*, vol. 2 : *L'analyse des correspondances*, Paris, Dunod.

BRUCKER F., BARTHÉLEMY J.-P. (2007), *Éléments de classification, aspects combinatoires et algorithmiques*, Paris, Hermès-Lavoisier.

ESCOFIER-CORDIER B. (1965), *Analyse des correspondances*, Thèse de doctorat, Faculté des sciences de Rennes, in *Les Cahiers du bureau universitaire de recherche opérationnelle*, 1969.

HAMMING R.W. (1950), "Error detecting and error correcting codes", *American Telephone and Telegraph Company* 26(2), p. 147-160.

JACCARD P. (1912), "The distribution of the Flora in the Alpine Zone", *New Phytol* 11, p. 37-50.

KALA KAMDJOUG J.-R., LENCA P., BARTHÉLEMY J.-P. (2007), "Assessment of actions in a multi-actor and multicriteria framework: application to the refunding of microfinance institutions", *Computational Economics* 29(2), p. 213-227.

KERBOAL M., BANSARD J.-Y., COATRIEUX J.-L. (2006), "An analysis of IEEE publication", *IEEE Engineering in Medicine and Biology Magazine* 25, p. 6-9.

LERMAN I.C. (1970), *Les bases de la classification automatique*, Paris, Gauthier-Villars.

LUONG X. (1988), *Méthodes d'analyse arborée. Applications*, Thèse de doctorat, Université Paris V.

PAGE R.D.M. (1996), "TREEVIEW: an application to display phylogenetic trees on personal computers", *Computer Applications in the Biosciences* 12, p. 357-358.

SAITOU N., NEI M. (1987), "The neighbor-joining method: a new method for reconstructing phylogenetic trees", *Mol. Biol. Evol.* 4, p. 406-425.

SATTAH S., TVERSKY A. (1977), "Additive similarity trees", *Psychometrika* 42, p. 319-345.