Nouvelles fonctionnalités pour la visualisation des données textuelles et des résultats : Pour une approche ergonomique des dispositifs lexicométriques

Jean-Marc Leblanc

Université Paris-Est Créteil Val de Marne, Centre d'Etude des Images ;
Textes, Ecrits, Communications (Céditec EA 3119)

Résumé

Un certain nombre d'initiatives vise aujourd'hui à comparer les logiciels et à mettre en commun les pratiques et les fonctionnalités. Beaucoup d'études et d'analyses mobilisent deux voire trois outils simultanément, dans le but de porter des éclairages différents sur des corpus identiques. Si certaines mesures tendent à se développer dans la plupart des outils (comme les cooccurrences, visualisables désormais aisément sous forme de graphes), il semble important de souligner certaines complémentarités entre les outils contrastifs ou longitudinaux et ceux qui produisent des cartographies globales sans nécessité de partitions intrinsèques. Il semble par ailleurs nécessaire aujourd'hui de se pencher sur des considérations ergonomiques, souvent peu représentées dans les dispositifs que nous connaissons. Enfin, nous formulerons quelques propositions pour de nouvelles fonctionnalités visant à faciliter la visualisation des données textuelles et des résultats. (Implémentation du mouvement dans les analyses multidimensionnelles, prise en compte du multimédia, navigation au sein des analyses factorielles…)

Abstract

Comparing and sharing software capabilities and functionalities and the way they are used is on the way. Many academics works use two or three software at the same time in order to extract different point of view about the same corpus. Even if some tools can be found in most of the software (coocurencies pictured with graphics) it seems to be important to highlight complementarities between contrasting tools, longitudinal tools and those which create global mappings without needing creating partitions. Besides, it seems to be necessary to work more on ergonomics facts. At last we will submit sketches of new functionalities in order to improve visualisation of textual data and results (movement in AFC, using multimedia tools, 3D navigation in AFC…).

Keywords: functionalities, ergonomics, visualisation, textometry, lexicometry

1. Comparer les logiciels aujourd'hui, cela a-t-il encore un sens?

Les fonctionnalités des outils textométriques évoluent et tendent à se recouper, d'une part dans une volonté de mise en commun, d'autre part en raison des possibilités qu'offre désormais la technologie. Ces fonctionnalités présentent souvent de fortes similitudes dans leurs principes (sur les cooccurrences, Leblanc, 2004), mais il convient de noter la complémentarité des outils contrastifs et longitudinaux d'une part (Lexico, Hyperbase, Weblex) et ceux qui permettent, plutôt par le biais de cooccurrences généralisée (Alceste, Astartext dans un certain nombre des ses fonctionnalités et désormais Hyperbase pour ce qui concerne le calcul des corrélats), de faire émerger des structures signifiantes indépendantes de partitions pré établies [1].

[1] Nous ne pouvons évoquer ici que quelques exemples de logiciels que nous utilisons et connaissons de façon

Les facilités technologiques, et un mouvement évident de mise en commun, engagent les concepteurs à grouper au sein d'un même outil les fonctionnalités disponibles ici ou là, oubliant parfois les postulats méthodologiques qui sont à l'origine de tel ou tel logiciel. Ainsi, chaque outil propose désormais, outre le calcul des cooccurrences, des réseaux d'attirances (lexicogrammes récursifs, pôles, corrélats), chacun proposant une variante de l'algorithme et une visualisation mais, tout ceci ne revient il pas, finalement au même [2] ? Si la complémentarité est évidente dans certains cas (exemple d'une analyse en tri croisé sous Alceste), ne devons-nous pas concentrer notre attention sur d'autres champs ?

2. Intégrer des préoccupations ergonomiques au sein des implémentations du domaine

Nous suggérons d'adopter aujourd'hui une approche plus ergonomique, que nous ne l'avons fait jusqu'ici, notion qui ne relève pas d'une simple considération d'ordre esthétique, on le sait. L'évolution des pratiques (internet, blogs et autres dispositifs comme itunes, daylimotion ou youtube et réseaux sociaux) a familiarisé les utilisateurs à des interfaces évoluées. L'ergonomie on le sait a d'abord avant tout servi les intérêts d'industriels qui souhaitaient rentabiliser le travail des ouvriers, non pas en cherchant le bien être de ceux-ci mais en optimisant le nombre de gestes à effectuer en production à la chaîne. Cette notion a fini par s'assimiler à l'organisation des conditions de travail en vue d'améliorer le confort des travailleurs, par exemple en aménageant le mobilier de bureau ou d'autres éléments de l'environnement de travail. L'ergonomie dont il est question ici concerne le domaine de l'interaction homme machine. Nous y incluons l'accès rapide et aisé aux données, mais aussi leur visualisation sous une forme qui fasse sens [3]. Cette notion de visualisation des données textuelles et des résultats serait facilitée par l'introduction du multimédia dans nos outils.

Les spécialistes du multimédia et de l'interaction homme machine, mais aussi les concepteurs de sites web sont depuis longtemps familiers des exigences ergonomiques. Outre les travaux de Nielsen et Molich (1990), de Boucher (2007) et de quelques autres, Bastien et Scapin (1993 ; 1998 ; 2001) proposent un certain nombre de critères fondateurs qui ne s'appliquent pas uniquement au web mais bien à tout dispositif informatique. Si nous considérons que les outils que nous utilisons au sein de la communauté scientifique sont destinés à des spécialistes il est évident que ces critères sont d'une nécessité moins impérieuse que s'il s'était agi de s'adresser à un plus large public. Il n'en reste pas moins que réfléchir à ces exigences peut nous permettre de rendre meilleures nos interfaces, et de les adapter à un public auquel il convient de penser également, ceux qui se forment à ces outils. La formation à la recherche est en effet un élément essentiel de notre démarche. Enfin, nous ne saurions occulter le fait que les moyens dont dispose la recherche publique ne permettent souvent pas à nos concepteurs d'intégrer

approfondie. Nous n'ignorons pas pour autant des initiatives telles que la plateforme textométrique (Heiden et Pincemin et al.), les développements récents de Lexico 3 (Salem), ou encore les travaux menés autour du trameur (Fleury). Ne pouvons ni fonder notre expertise sur des outils qui n'existent pas encore à l'heure où nous terminons cette contribution, ni évoquer des logiciels dont nous n'avons pas une maîtrise suffisante. En aucun cas nous ne prétendons ici à l'exhaustivité. Taltac (Bolasco), Sato (Daoust), et bien d'autre pourraient faire l'objet d'une analyse, mais là n'est pas notre propos.

[2] Affirmation provocatrice qui introduit à d'autres travaux où nous proposons une comparaison de résultats issus de quelques outils : classes thématico-sémantiques d'Alceste, analyses factorielles de cooccurrences généralisées, analyses de formes issues de tableaux lexicaux.

[3] Sur les problèmes de visualisation voir aussi le numéro spécial de la revue Transeo, *figurer l'espace en sciences sociales*. http://www.transeo-review.eu.

toutes ces dimensions qui nécessitent des heures ingénieur non négligeables mais aussi des tests utilisateurs et autres protocoles qui sont plus largement utilisés dans le secteur privé dont les objectifs et exigences diffèrent par ailleurs sensiblement. Cela ne signifie pas pour autant que les logiciels dits commerciaux soient parfaitement ergonomiques. Tropes, Sphinx, ou encore Alceste ont quelques progrès à faire dans le domaine, ainsi que nous le verrons.

Les critères de Bastien et Scapin, sont fournis ici à titre d'exemple mais mériteraient d'être adaptés à notre domaine particulier sous forme d'une nouvelle grille de critères ergonomiques destinée à la textométrie/lexicométrie.

Si l'outil est en ligne, il doit satisfaire à d'autres exigences encore, celles qui s'appliquent à tout site web. Nous ne détaillerons pas ici mais renvoyons aux références précédemment citées.

2.1. Principaux concepts ergonomiques (Bastien - Scapin)

- Guidage : incitation, groupement/distinction entre items (par le format, par la localisation), feed back immédiat, lisibilité ;
- Charge de travail : brièveté (concision, actions minimales), densité informationnelle ;
- Contrôle explicite : actions explicites, contrôle utilisateur ;
- Adaptabilité : flexibilité, prise en compte de l'expérience de l'utilisateur ;
- Gestion des erreurs : protection contre les erreurs, qualité des messages d'erreur, correction des erreurs ;
- Homogénéité/cohérence ;
- Signifiance des codes et dénominations ;
- Compatibilité.

2.1.1. Le concept de guidage

Le concept de guidage concerne l'ensemble des moyens mis en œuvre pour « guider » l'utilisateur, l'informer l'orienter. Cette notion comprend un certain nombre d'éléments que nous ne pourrons développer ici. L'incitation en est un des sous critères important. Pris dans une acception relativement large, il désigne, outre le fait d'amener l'utilisateur à exécuter une action, tous les messages permettant de connaître les alternatives disponibles lorsque plusieurs actions sont possibles, toutes les informations permettant à l'utilisateur de savoir où il en est d'un cheminement, d'une opération. En cela Hyperbase [4] répond assez bien à ce principe en indiquant les différents étapes du traitement d'un corpus (données lemmatisées, segmentation, calculs de proximités) et le temps restant. La documentation d'un outil est donc primordiale [5]. Sous Lexico, lorsque l'on active l'icône « principales caractéristiques d'une partition », le menu vertical peut être considéré comme une manière d'indiquer les alternatives possibles [6] (spécificités, AFC…).

[4] Notons la difficulté qu'il y a à localiser les fichiers texte sous Hyperbase. (Lors de la création de nouveaux corpus) en raison composition des noms fichiers.

[5] Tropes que nous citons peu ici est remarquablement documenté.

[6] L'espace manque dans un tel format pour fournir toutes les illustrations des cas cités ainsi que les propositions de fonctionnalités à développer. Ces éléments, figures et maquettes seront disponibles dès la tenue du colloque à l'adresse http://textopol.org/contrib2010.php.

Conformément au principe d'*incitation* et de *guidage* on évitera des pages trop longues comme on peut en trouver lorsqu'il s'agit de retourner au texte, en proposant des systèmes de navigation, d'ancres, d'accès directs qui facilitent la lecture. Ce qui est vrai d'un site web peut l'être de nos outils : une proportion très importante d'utilisateurs préfère une lecture papier qu'une lecture écran. Il est donc nécessaire d'améliorer cette lecture sur nos dispositifs. On proposera des adressages, et ce n'est pas là une question ergonomique mais bien une préoccupation méthodologique : quelle est la partie du corpus concernée, quel paragraphe, à quelle page, à quelle date, quel est le locuteur concerné, autant d'indications indispensables à l'interprétation. La fonction *référence* de Weblex est en cela intéressante. En intégrant un nombre suffisant de balises sous lexico 3, on peut contourner ce que nous pensons être une insuffisance – relative – de ce point de vue (une balise date, page, un caractère matérialisant les paragraphes, que l'on choisirait ensuite pour le « regroupement » des contextes par exemple).

Sur un autre plan, il est utile également, toujours dans le cadre de ce critère, d'indiquer le temps que peut prendre une action (sur le web on indiquerait le poids d'un document disponible au téléchargement), ici on pourrait annoncer le temps de traitement d'un corpus ou d'un calcul particulier. Hyperbase d'ailleurs répond fort bien à ce principe. Ceci ne reste valable que dans le cas de longs traitements. Or, il convient de noter que dans la plupart des cas les calculs sont exécutés de façon quasi instantanée. Hyperbase et Alceste sont les deux outils, parmi ceux que nous utilisons, qui proposent des traitements initiaux assez longs, sans doute parce que tous les traitements sont exécutés par défaut, alors que d'autres outils ne réalisent que la segmentation et la construction du tableau lexical, les autres calculs étant effectués ensuite.

Groupement/distinction entre les items : qu'il s'agisse de résultats ou de fonctionnalités, on groupera les items selon les catégories auxquelles ils appartiennent, en les distinguant, soit sur le plan topologique, soit sur celui du visuel (couleur ou polices). Nous citerons plus bas le cas de Weblex et des cadres qui remplissent cet office et le cas d'Alceste (version 4.8) qui présente sur un même plan des informations d'ordres très différents, qu'il s'agisse de résultats ou de fonctionnalités. (Nombre de classes, nombre d'énoncés, proportion d'énoncés classés catégories grammaticales…). Notons par ailleurs la grande difficulté qu'il y a à sauvegarder la configuration des classes avec cet outil. Ceci nous amènerait à aborder la question du format des résultats et de leur exportation, mais aussi des options qui pourraient être proposées en fonction de la destination de ces visualisations (impression, écran, exports, enregistrement des tableaux, vectoriel plutôt que bitmap, données tabulées…).

Dans la représentation des résultats lexico 3 opte pour un histogramme lorsqu'il s'agit de spécificités, et pour des courbes s'agissant de fréquences relatives ou absolues. Ce choix entre en contradiction avec une exigence sémiotique qui voudrait qu'on représente des données discrètes par des histogrammes et des données longitudinales par des courbes [7]. Il pourrait être intéressant que l'utilisateur puisse choisir les options des graphiques (types, fonds, couleurs, légendes). La distinction entre spécificités et fréquence ne serait plus véhiculée par le type de graphique (histogramme ou courbe) mais bien par d'autres moyens.

Autre cas de figure, distinguer les fonctionnalités qui sont disponibles de celles qui ne sont pas activables, toujours par des visuels différents. Pas d'AFC à moins de trois parties, pas de lexicogramme sans forme pôle… De la même façon il est essentiel d'éviter ce que nous faisons tous, proposer dans un menu des fonctionnalités qui ne sont pas encore implémentées et donc disponibles.

[7] Hyperbase propose beaucoup d'histogrammes également.

Sur les outils qui s'y prêtent, c'est-à-dire qui présentent une navigation linéaire ou par étapes, mettre en évidence l'icône qui correspond à l'écran ou à l'étape à laquelle on se trouve.

Nous passerons rapidement sur le concept de Feedback immédiat évoqué au sein d'autres étapes mais il convient de souligner que toute action doit avoir un résultat (message, réponse, tableaux), et ce, dans un délai relativement court.

2.1.2. Charge de travail

La *brièveté* et surtout la notion *d'action minimale*, imposent de limiter le nombre des étapes pour l'accomplissement d'une tâche. En cela l'approche de Lexico 3 et des « glisser/déplacer » est remarquable.

Selon le critère de concision, on limitera le nombre d'items (menus ou résultats) à 7 ou 9 par écran afin d'éviter tout surcharge cognitive. Sous Lexico 3 on pourrait souligner que lorsque l'utilisateur a effectué un certain nombre de mesures, il se trouve envahi de fenêtres que l'on peut déplacer sur une autre feuille mais dont il est difficile de se débarrasser simplement. Cela concerne aussi la notion de densité informationnelle. Selon ce principe il convient de ne pas surcharger l'utilisateur quant aux choix de liens qu'il peut activer…. Si ce concept évoqué par Bastien et Scapin, se rapporte plutôt aux dispositifs en ligne, nous étendrons cet exemple au cas du logiciel Hyperbase. Nous suggérons que la multiplicité des visualisations, et des options de calcul peut entraver ce principe [8].

L'outil doit également exécuter les seules actions explicitement initiées par l'utilisateur. C'est ce que les auteurs nomment le *contrôle explicite*. Il s'agirait par exemple de proposer des boutons ou des formulaires de validation. Cela n'aurait de sens que pour des traitements relativement longs. Selon le même principe l'utilisateur doit avoir tout loisir d'interrompre un traitement, mais aussi de revenir en arrière. Lexico 3 propose d'effacer une partie du contenu d'une carte des sections. En revanche sur les graphiques on aimerait pouvoir n'effacer que la dernière forme glissée et non tout l'ensemble. Cette possibilité d'annuler la dernière action, mais aussi de la répéter si besoin, correspondrait au CTRL Z ou CTRL Y de la plupart des logiciels courants. Enfin, selon ce principe, il devrait toujours être possible d'accéder au menu initial, ce que permet Lexico 3 de par sa configuration même (une feuille unique tout au long des navigations) mais aussi Hyperbase. Dans ce domaine, il est bon par ailleurs que l'utilisateur ait clairement conscience de la nature des calculs mobilisés et des méthodes employées, des principes. Les « boites noires » sont donc à proscrire.

2.1.3. Adaptabilité

Bien qu'entrant en contradiction avec ce que nous notions précédemment s'agissant de la multiplicité des options disponibles sous Hyperbase, ces choix peuvent être une manière de répondre à l'exigence d'adaptabilité : prendre en compte l'expérience de l'utilisateur, guider

[8] La multiplicité des modes de calcul et des analyses multidimensionnelles fournies par *Hyperbase* suppose une connaissance approfondie en statistique générale. Calcul de la distance lexicale, sur les formes ou sur les occurrences, factorielle sur V (vocabulaire, formes) ou sur N (occurrences), analyse arborée, sur N, sur V, portant, au choix sur les distances brutes, sur le CHI2 ou sur le carré des distances, en présentation radiale ou rectangulaire, mais aussi histogrammes lignes colonnes, portant sur les formes ou les occurrences, l'histogramme de la richesse lexicale portant sur les hapax ou sur le vocabulaire sans évoquer les analyses possibles sur les annotations du corpus, accroissement lexical par tranches ou portant sur les divisions naturelles du texte, tout ceci rend le logiciel *Hyperbase* peu accessible au néophyte (Leblanc, 2005).

les novices et permettre d'autres choix aux experts. L'interface simplifiée de Weblex peut être une réponse à ce principe [9].

2.1.4. La gestion des erreurs et la protection contre les erreurs

On fera en sorte que nos outils fonctionnent quel que soit l'endroit où l'on place le corpus (AFC Lexico 3, installation Hyperbase, catégorisation Astartex…). La qualité des messages et la gestion des erreurs est probablement l'un des chantiers les plus importants du domaine : Pôle non défini sous Weblex (est-ce suffisant ?), erreur Baob2 (corpus trop court, pour Alceste [10]), erreur de lecture du fichier des facteurs sous Lexico 3, Error ; invalid pour Hyperbase et bien d'autres. Sur ce point, aucun logiciel ne satisfait au critère. Nos systèmes d'exploitation commerciaux voire monopolistiques ne sont pas toujours mieux lotis (écrans bleus, *no disk found*…) ce n'est que la notoriété, non pas des pannes mais des systèmes qui palie ce problème ergonomique.

2.1.5. Signifiance des codes et dénominations

Ceci est difficilement applicable à notre domaine où coexistent plusieurs terminologies. On évitera l'utilisation d'icones et de vocabulaire nécessitant un apprentissage particulier. Dans un domaine scientifique voilà qui n'est pas applicable mais le vocabulaire gagnerait à s'harmoniser d'une « Ecole » à l'autre, tout en respectant les postulats de départ de chacun (Forme, item lexical, les utilisateurs de Nooj utilisent une autre terminologie. Tout cela n'est pas sans poser problème à ceux qui doivent aborder ces méthodes) [11].

2.1.6. Compatibilité

Soulignons l'existence de normes d'accessibilité des sites web et des dispositifs multimédia. Nous ne traiterons pas cet aspect ici, bien qu'il soit primordial, car il ne s'applique pas à notre seul domaine mais bien à tous les dispositifs, quels qu'ils soient.

Si nous devions résumer ce qui précède, nous soulignerions que, bien que sa conception ne soit pas tout à fait récente, Hyperbase répond remarquablement bien à des exigences relatives aux concepts ergonomiques. Documentation des fonctions, annonce des temps de traitement, adressages de contextes, navigation dans le corpus, groupement des items, exports possibles de données pour tableurs, ponts avec d'autres outils sont autant d'atouts du logiciel. Son plus gros défaut est peut-être celui de ses qualités : son évolution rapide et incessante, implique que le manuel utilisateur n'est jamais à jour, que les calculs ne sont jamais vraiment les mêmes d'une version à l'autre. Il reste que pour lancer la segmentation d'une nouvelle base, il faut être tout de même assez connaisseur et qu'il ne s'agit pas d'une installation standard. On soulignera une certaine lourdeur dans les traitements qui font que cet outil est selon nous plus approprié pour les corpus qui ne sont pas destinées à évoluer fréquemment. Lexico 3 au contraire est l'outil de l'expertise rapide par excellence. Manipulations par glisser déposer, portabilité du logiciel (qui fonctionne sur une simple clé ce qui est remarquable), outil minimal (portant sur la forme graphique) Lexico permet par cette simplicité beaucoup plus que les autres. Il répond en cela remarquablement au critère de compatibilité.

[9] Il pourrait être intéressant de proposer un module qui permettrait de filtrer les types généralisés de Lexico 3, afin de tenir compte des emplois des mots en contexte.

[10] Toutes les erreurs Alceste sont codées de façon fort exotique, bien que le manuel les référencie toutes, mais pourquoi ne pas les expliciter ?

[11] Cette tentative d'harmonisation terminologique est une des préocupations des membres du projet Textométrie.

3. Proposition des fonctionnalités et de visualisations : cahier des charges

Cette réflexion nous amène à proposer un certain nombre d'implémentations destinées à développer la visualisation des données textuelles et des résultats en intégrant les dernières technologies multimedia. C'est un aspect du chantier que Textopol met en œuvre actuellement [12].

3.1. Intégrer une troisième dimension dans les analyses multidimensionnelles

Dans ses travaux sur les discours des Nations Unies sur le développement durable, Fabienne Pierre (2006) à souligné la difficulté de lecture d'une AFC sur deux axes dès lors que le nombre de parties est important. Elle propose alors, et expérimente, une représentation sur trois axes qui permet en effet une meilleure appréhension des phénomènes qu'elle cherche à analyser [13]. Nous pensons que cette réflexion peut être poursuivie au sein du chantier que nous mentionnions précédemment, tout en formulant une réserve relative. En effet, les analyses factorielles, destinées à représenter les phénomènes les plus saillants, ont-elles vocation à intégrer d'autres axes que les plus significatifs ? Pourquoi alors s'en tenir au troisième axe ? Malgré cette réserve, nous pensons cependant que les outils doivent intégrer la représentation tridimensionnelle, y compris dans les AFC [14].

3.2. Développer une navigation entre les différentes partitions d'un corpus

La deuxième proposition concerne la possibilité de naviguer entre les différentes partitions d'un corpus et la possibilité de zoomer sur les analyses factorielles dans une métaphore atomique du texte qui pourrait s'expliquer par l'imbrication des plusieurs AFC. Prenons un corpus de discours présidentiels partitionné par locuteur puis par année. La première AFC serait la représentation générique de la configuration par locuteur, puis zoomant sur un des locuteurs, cette AFC laisserait place à une configuration qui traduirait la distance entre les différentes années. Enfin, sur une année – donc un texte – on pourrait envisager de saisir la structure interne au moyen d'une approche cooccurrentielle. Cette proposition est née d'une collaboration avec Pierre portant sur le discours rituel.

3.3. Implémenter le mouvement dans les données issues d'analyses multidimensionnelles et longitudinales

La troisième proposition est issue de nos observations sur le discours présidentiel. Sur un corpus s'étendant de 1959 à 2001 les analyses factorielles pratiquées sous Lexico 3 et portant sur la partition année, montrent une singularité des textes du général de Gaulle, qui s'opposent à l'ensemble de ses successeurs. Cette singularité se réduit à mesure que de nouveaux textes sont introduits dans l'analyse. Lorsque nous ne disposions que des textes de 1959 à 2001, les messages du général de gaulle étaient beaucoup plus discriminés. A mesure que nous avons ajouté les textes suivants, de 2002 à 2008, nous avons vu les messages du premier président

[12] Cette exigence devient plus impérieuse à mesure que les applications web se développent et proposent déjà, des visualisations et des fonctionnalités très attrayantes, bien que privilégiant parfois la visualisation sur la pertinence des analyses.

[13] J.-M. Viprey a tenté lui aussi de résoudre cette difficulté de lecture en proposant notamment une projection géodésique du troisième axe dans son logiciel Astartex. Nous approfondirons cette réflexion sur la base des travaux de Viprey dans le cadre de contributions plus développées.

[14] Nous ajoutons par ailleurs que les AFC devraient comporter par défaut les informations liées à la contribution des différents points à la construction des axes.

se rapprocher du centre de l'AFC. Ceci nous conduit à proposer d'une part une nouvelle fonctionnalité destinée à représenter et à se mieux saisir des phénomènes progressifs de ce type en proposant d'implémenter le mouvement dans certaines visualisations lexicométriques et à ébaucher d'autre part une hypothèse : Chirac et Sarkozy tentent-ils, sur la dernière période de notre corpus, d'adopter une posture gaullienne ?

Plus généralement, nous proposons d'implémenter le mouvement au sein de toute visualisation qui s'y prête afin de mettre en évidence les variations de paramètres, de seuils, d'élagages, qu'il s'agisse de calculs de cooccurrences représentés sous forme de graphes ou de réseaux ou encore de mesures longitudinales. Ces différentes propositions seront disponibles en ligne à partir de la tenue du colloque et visualisables sur place sous formes d'animations et de poster.

4. Conclusion

A l'issue de ces réflexions encore inachevées, une première conclusion s'impose. Il est peut-être temps que des chercheurs, travaillant parfois dans les mêmes universités, se rencontrent et mettent en commun leurs compétences. Spécialistes du multimédia, des interfaces, de l'hypertextualité, de l'édition numérique, de l'ergonomie, de la visualisation, et du domaine qui nous intéresse ici, celui de la lexicométrie, textométrie, logométrie…C'est le projet que l'équipe Textopol du Céditec voudrait initier en réunissant ces compétences autour de la rédaction d'un cahier des charges pour la visualisation des données textuelles, projet qui prendrait en considération les préoccupations liées à l'ergonomie.

Très attaché à la formation à la recherche, nous estimons, malgré un titre de première partie un peu provocateur, qu'il est important que différents outils existent, qui sont le reflet des postulats, d'écoles, de préoccupations, de formations mais aussi d'objets d'étude différents. Il reste que nous avons des progrès à faire autant sur le plan de l'organisation des interfaces, et de la navigation, que dans la visualisation des résultats, car le web offre désormais un éventail de fonctionnalités et de représentation dont nous devons tenir compte. Nuançons cependant notre propos pour terminer en soulignant que la visualisation ne doit pas prendre le pas sur le contenu…

Références

Bastien J.M.C. and Scapin D.L. (1993). Ergonomic criteria for the evaluation of human-computer interfaces. Technical report Ndeg. 156. INRIA.

Bastien, J. M. C. and Scapin, D. L. (2001). Évaluation des systèmes d'information et Critères Ergonomiques. In C. Kolski (Ed.), *Systèmes d'information et interactions homme-machine. Environnement évolués et évaluation de l'IHM. Interaction homme-machine pour les SI* (Vol. 2, pp. 53-79). Paris : Hermes.

Bastien J.M.C. Leulier C. and Scapin D.L. (1998). *L'ergonomie des sites web.* INRIA Rocquencourt.

Boucher A. (2007) *Ergonomie web*. Eyrolles.

Christiane P. (2004). L'art numérique. Chapitre : bases de données, visualisation de données et cartographie. Paris : Thames et Hudson.

Leblanc J.M. (2004). Développer la méthode expérimentale en analyse textuelle, *Communication au Colloque International sur le Document Électronique (CIDE7, La Rochelle, 22-25 juin 2004)*.

Leblanc, J.M. (2005). *Les vœux des présidents de la cinquième République (1959-2001). Recherches et expérimentations lexicométriques à propos de l'ethos dans un genre discursif rituel*, Thèse de

Doctorat en Sciences du Langage, Créteil : Université de Paris 12 Val-de-Marne, 2005. (Sous la direction de Pierre Fiala).

Nielsen J. and Molich R. (1990). Heuristic evaluation of user interfaces In *Proceedings ACM CHI'90 Conf.*, Seattle, WA, 1-5 April, pp. 249-256.

Pérès M. (2001). *Réflexions sur le modèle informatique du Circus Maximus*. Thèse de doctorat, Histoire, Langue et littérature anciennes, Université Michel de Montaigne Bordeaux3, 2001.

Pierre F. (2007). *Espace public transnational. Eléments épistémologiques et méthodologiques pour l'analyse. A travers le Sommet Mondial du Développement Durable (Johannesburg 2002)*. Thèse de doctorat, sciences de l'information et de la communication, dir. Simone Bonnafous, soutenue en décembre 2007.

Roxin I., Hutschmitt B., Mercier D. and Leblanc J.M. (2007) Web sémantique, navigation dans de grands corpus textuels In *CIDE 10, dixième colloque international sur le document numérique*, 2-4 juillet 2007, Nancy, INIST.

Scapin D.L. (1990). Des critères ergonomiques pour l'évaluation et la conception d'interfaces utilisateurs. In *Actes du XXVI° Congrès de la SELF*, Montréal, Canada. - Institut de recherche en Santé et Sécurité au Travail du Québec.

Simonin J. and Carbonell N. (2007). Interfaces adaptatives : adaptation dynamique à l'utilisateur courant. In Saleh, I. and Regottaz, D., *Interfaces numériques*, Paris : Hermes Lavoisier (coll. Information, hypermédias et communication).

Villey-Migraine M. (2004). Evaluation ergonomique de produits multimédias pour des publics spécialisés: Apports méthodologiques et sélection des mesures en fonction du contexte. In *Proceeding of Informations, Savoirs, Décisions et Médiations* (ISDM, Toulon-Marseille, France, 12 January 2004).

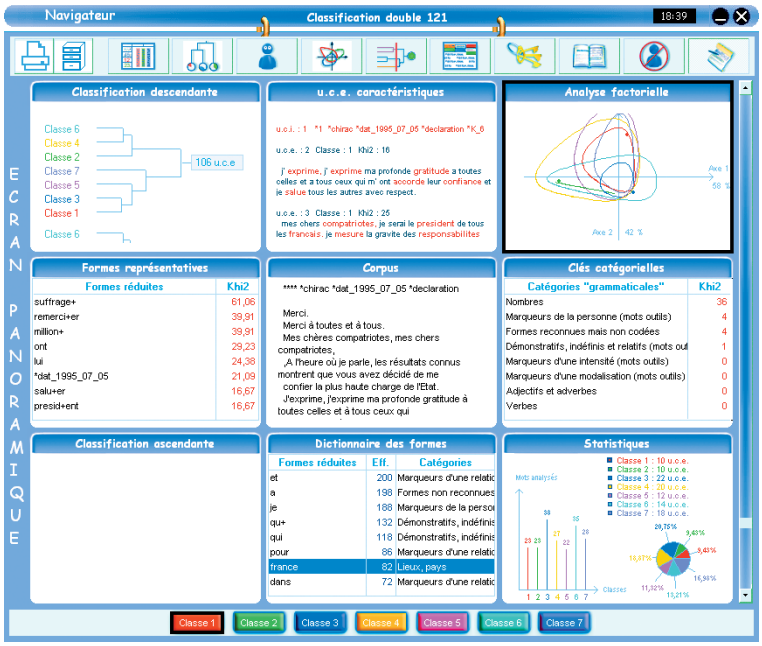
Viprey J.-M. (2006), Ergonomiser la visualisation AFC dans un environnement d'exploration textuelle : une projection « géodésique ». In *JADT 2006*.

Lergonome - http://www.lergonome.org/.

Ergolab - http://www.ergolab.net/.

ANNEXES

Réflexions et préconisations :



Exemple d'Alceste…

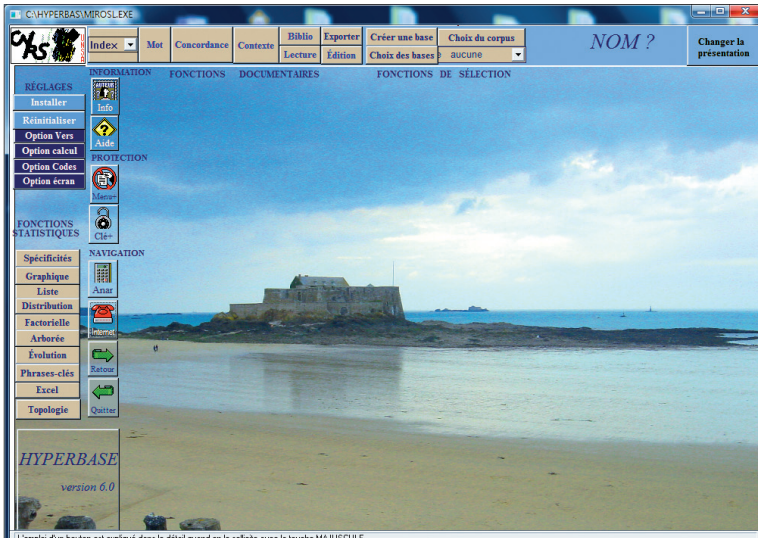
Cet écran présente des informations de niveaux différents.

Résultats statistiques, corpus, retour au texte (uce caractéristiques) et dictionnaire des formes…

Les différents boutons du haut de l'écran sont compréhensibles du moment que l'on connaît la visualisation des différentes représentations…

L'interface d'Hyperbase

Respecte d'une certaine manière le concept de distinction/groupement entre les items, que ce soit topologique ou graphique, en distinguant en haut du panneau les actions relatives aux fonctions documentaires de celles qui concernent des mesures. Un troisième groupe de boutons est même signalé par un code couleur différent et par une zone distincte des autres pour les options de calcul. Le nombre des items reste très important, mais peut on vraiment le limiter ?



La présentation linéaire de l'interface de Lexico 3, est assez intéressante. Les info-bulles sont appréciables. Sur certaines versions elles ne sont malheureusement pas disponibles partout (exemple sur les cartes des sections et les spécificités appliquées aux sections).

On pourrait imaginer que le fond de cette barre des menus soit d'une couleur différente pour les fonctionnalités de traitement d'un corpus (ouverture et segmentation) pour les fonctions plus documentaires et pour les calculs à proprement parler. On imaginerait assez bien aussi que les principales caractéristiques des partitions soient placées parmi les premiers boutons. Quant aux icones utilisées, elles sont assez efficaces pour qui est familier de l'outil, mais pas toutes très explicites sur le plan de leur signification. Celles qui concernent la carte des sections ainsi que nouvelle feuille, enregistrement au rapport, concordance, nouvelle base, ouverture de dossier sont très efficaces de ce point de vue, mais d'autres sont moins évidentes.

L'interface de Weblex sépare en zones bien distinctes ce qui concerne l'affichage des résultats, la saisie des données (motifs, sources, recherche, pôles, ou focus), les fonctionnalités à proprement parler (entre parenthèses lorsqu'elles s'appliquent à un pôle) et les paramètres. Des couleurs différentes permettent de distinguer plus aisément ces classes différentes. L'interface reste très complexe et très riche mais la simplification semble peu envisageable quand on connaît la richesse de l'outil et les possibilités qui sont offertes. Même chose pour le langage d'interrogation (CQP) qui n'est pas du tout à la portée du non initié. On peut noter une tentative de lexico 3 en ce sens, mais aussi d'Hyperbase de proposer des traductions de ces expressions régulières.

On notera quelques ambigüités (selon nous) qui ne peuvent être levées que si on est rompu à la pratique de weblex : la distinction entre le singulier et le pluriel pour les *lexicogrammes simples* ou *récursifs*, la dénomination de *lexicogramme* qui s'applique à deux présentations forts différentes (récursifs ou simples) mais aussi spécificités total alph, total fréquence, total parties…