



Développement d'un outil de cartographie automatique de texte d'aide à la compréhension de texte pour des apprenants dyslexiques

Mario Laurent, Thierry Chanier

► To cite this version:

Mario Laurent, Thierry Chanier. Développement d'un outil de cartographie automatique de texte d'aide à la compréhension de texte pour des apprenants dyslexiques. Environnement Informatique pour l'Apprentissage Humain 2013, May 2013, Toulouse, France. edutice-00854240

HAL Id: edutice-00854240

<https://edutice.archives-ouvertes.fr/edutice-00854240>

Submitted on 26 Aug 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Référence pour cité la communication :

Laurent, M., Chanier, T. (2013). Développement d'un outil de cartographie automatique de texte d'aide à la compréhension de texte pour des apprenants dyslexiques. *Poster conférence EIAH (Environnement Informatique pour l'Apprentissage Humain) 2013*, 29-31 mai, Toulouse.

Texte intégral de la proposition de communication, tel qu'il a été validé :

Outil d'aide à la compréhension de texte pour des apprenants dyslexiques

Développement d'un outil de cartographie automatique de texte d'aide à la compréhension de texte pour des apprenants dyslexiques

Mario Laurent , Thierry Chanier***

** Laboratoire de Recherche sur le Langage,
Université Blaise Pascal, Maison des Sciences de l'Homme
4, rue Ledru
63057 Clermont-Ferrand cedex 01
thierry.chanier@univ-bpclermont.fr*

*** Laboratoire de Recherche sur le Langage,
Médialexie
6 rue Nicolas Joseph Cugnot
63100 Clermont-Ferrand
mario.laurent@etudiant.univ-bpclermont.fr*

RÉSUMÉ. Les personnes atteintes de troubles du langage et notamment la dyslexie éprouvent de grandes difficultés à lire et à écrire. Ces problèmes persistent lorsqu'il s'agit d'utiliser un environnement informatique, il est donc nécessaire de leur fournir un ensemble d'outils adaptés. Nous proposons de réaliser un outil, le LICI, permettant d'obtenir automatiquement une carte conceptuelle ou heuristique à partir d'un texte afin d'aider ces personnes à comprendre un texte en un temps réduit et ainsi leur permettre de gagner du temps au cours d'activités de documentation ou d'apprentissage.

MOTS-CLÉS : Traitement Automatique des Langues ; dyslexie ; résumé automatique ; remédiation ; environnement d'apprentissage personnalisé

1. Contexte

1.0. La dyslexie

L'étude de la littérature scientifique abordant les troubles du langage et en particulier la dyslexie développementale permet de constater que ce trouble est complexe et multiple. Un pourcentage important de la population d'une classe d'âge est atteint et chaque cas est différent [INSERM (dir.) 07]. En effet, le diagnostique est différentiel et manque de normalisation, au final de nombreux cas différents sont regroupés derrière l'étiquette dyslexie [JACQUIER-ROUX et al. 10]. Ajoutons à cela qu'une part importante des dyslexiques souffre également d'un ou plusieurs autres troubles, notamment la dysorthographe et la dyscalculie [HABIB & JOLY-POTTUZ 08]. Ces personnes sont donc en réelle difficulté pour lire et produire des documents écrits. Ces raisons et l'omniprésence des environnements informatiques dans les cursus d'études nous poussent à nous poser la question de l'adaptation de ces environnements pour les personnes atteintes de troubles du langage qui sont en phase d'apprentissage. Afin de leur permettre une intégration sociale mais principalement et d'abord de favoriser leur cursus scolaire, il est essentiel de mettre à leur disposition des outils d'aide et de remédiation.

1.1. Outils d'aide et de remédiation

Depuis le début des années 2000, de nombreux outils développés précédemment en TAL ont été adaptés dans le but d'apporter une aide aux personnes atteintes de troubles du langage. On peut notamment citer des outils de : synthèse vocale, reconnaissance vocale (utilisée pour piloter l'ordinateur ou pour rédiger un texte), correction automatique, prédiction de mots, et enfin, de conception de cartes conceptuelles ou heuristiques. L'efficacité de ces outils a été validée par plusieurs études [HECKER & ENGSTROM 07]. L'entreprise Médialexie, à l'origine de nos travaux, propose déjà une barre d'outils rassemblant l'ensemble de ces fonctionnalités [MÉDIALEXIE 12]. Notamment un logiciel permettant de dessiner des cartes heuristiques adapté à son public. Nous souhaitons maintenant développer un outil complémentaire de transformation automatique d'un texte en une carte conceptuelle ou heuristique, que nous appellerons ici simplement carte. [MARGARIDO et al. 08] ainsi que [NANDHINI & BALASUNDARAM 11] ont montré que les résumés automatiques peuvent apporter une aide en ce qui concerne la compréhension pour les lecteurs en difficultés. L'idée est d'aller un peu plus loin et de remplacer les phrases produites en sortie d'un programme de résumé automatique par une carte qui sera d'autant plus lisible par le public dyslexique.

2. Cartographie automatique : le LICI

2.0. Tâches envisagées

Une lecture suffisamment fluide est indispensable pour réaliser l'intégration sémantique d'un texte. Cette intégration correspond au passage de l'état linéaire du texte à une représentation mentale des concepts sous-jacents et des relations entre les différentes entités évoquées [COIRIER et al. 96]. Grâce à la cartographie automatique, nous espérons compenser la difficulté d'intégration sémantique des personnes dyslexiques en leur

proposant directement une représentation graphique de ces concepts et relations. L'outil développé, que nous appellerons LICI, pour Logiciel Intelligent de Cartographie Inversée, s'adresse d'abord à un public d'élèves de collège/lycée. Ces élèves sont confrontés régulièrement à la lecture de textes relativement longs, ce qui multiplie leurs chances d'erreurs et risque de les mener à une mauvaise compréhension du texte. Dans sa version finale, le LICI leur permettra d'importer un fichier texte ou d'utiliser un simple copier/coller pour en obtenir automatiquement une version schématique. Cette carte présentera les éléments jugés les plus importants pour accéder au sens du texte et des liaisons entre ceux-ci, représentant leur relations. Les éléments seront des mots ou des groupes de mots, comme des groupes nominaux. Les relations pourront être d'ordre sémantique, spatial ou encore chronologique, selon le type de texte traité. Cet outil permettra, par exemple, de gagner du temps lors d'une activité de documentation au cours de laquelle un apprenant a besoin de lire plusieurs textes trouvés en ligne et souhaite sélectionner le plus pertinent pour son travail. Il sera également très utile en situation d'apprentissage d'une leçon en faisant ressortir immédiatement les notions les plus importantes.

2.1. Fonctionnalités

Un soin particulier sera accordé à la réalisation de l'interface, elle sera simple afin de ne pas substituer la difficulté d'utilisation à la difficulté de lecture. Les éléments affichés sur la carte produite seront accessibles à l'aide de la synthèse vocale. C'est-à-dire que la carte pourra s'afficher au fur et à mesure, chaque élément étant lu par une voix de synthèse, ce qui permettra de conforter l'utilisateur dans sa compréhension. Le profil de l'utilisateur sera dynamique et l'élève pourra l'enrichir selon ses spécificités au fur et à mesure de l'utilisation du LICI. Il permettra de sauvegarder des préférences, notamment les choix de présentation de la carte, comme le niveau de précision, l'espacement entre les caractères et entre les mots ou encore les couleurs du fond et des caractères. Par niveau de précision nous entendons le nombre d'éléments (nœuds) affichés et leurs plus ou moins grandes spécificités. L'espacement entre les caractères et entre les mots est très important, [ZORZI et al. 12] ont montré qu'une simple augmentation de l'espace entre les caractères augmente de façon significative les performances des dyslexiques. Enfin, à partir de chaque élément de la carte, on pourra accéder soit à un aperçu du texte présentant chacune de ses occurrences surlignées, soit à un concordancier. Un lien vers des dictionnaires sera également fourni. La carte sera de plus éditable pour permettre aux utilisateurs de la retravailler.

2.2. Traitement automatique

Pour réaliser l'outil de cartographie automatique, nous allons nous appuyer sur des technologies du domaine TAL déjà développées dans d'autres objectifs. Par exemple les techniques de résumé automatique de texte [TORRES-MORENO 11]. Concernant la chaîne de traitement, nous envisageons de procéder de façon classique, en nous inspirant du processus standard pour obtenir un résumé automatique, représenté sur la figure 1.

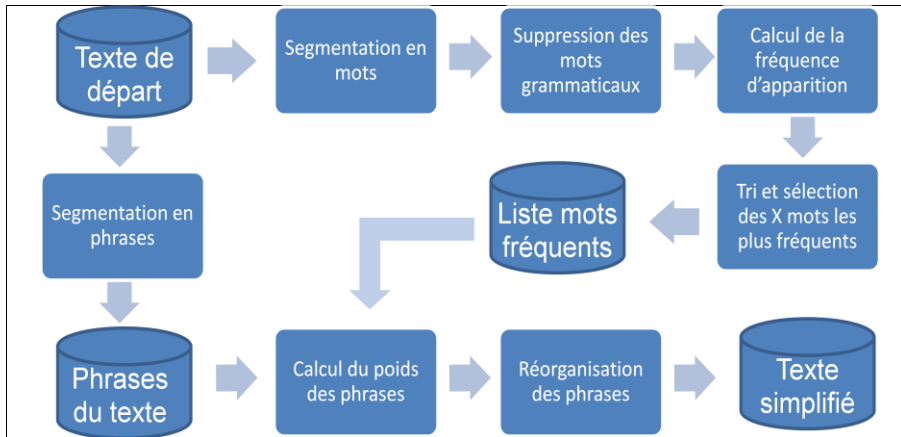


Figure 1. Chaîne de traitement pour le résumé automatique de texte par extraction de phrases

Nous n'avons cependant pas à nous soucier des phrases. Nous effectuerons d'abord un prétraitement : segmentation en graphies, puis étiquetage des graphies avec un outil du type [TREETAGGER 12] pour obtenir leurs catégories grammaticales et leurs lemmes. Nous pourrons ensuite effectuer un tri selon les catégories grammaticales des graphies. Ensuite, nous réaliserons le comptage de fréquence des graphies, regroupées par lemme. Nous tiendrons compte, en plus de la fréquence d'apparition, de la position des éléments dans le texte : les éléments apparaissant dans le titre, le premier et le dernier paragraphe pouvant être davantage significatif dans le cas d'un article par exemple. A ce stade il est capital de déterminer si des éléments fréquents sont co-occurents pour construire les futurs éléments de la carte qui peuvent être constitué de plusieurs graphies. Ainsi le programme devra reconstruire les groupes nominaux, par exemple « chat noir » ou du « Général De Gaulle » et ne pas traiter les graphies qui les composent de façon isolées. Enfin, nous pourrons faire appelle à une ressource extérieur du type arbre ontologique sémantique, pour remplacer un ensemble d'éléments par leur hyperonyme lorsque c'est pertinent, ou pour adapter la précision de la carte aux demandes de l'utilisateur. Une autre possibilité de traitement serait de s'appuyer sur les méthodes de fouilles de textes permettant la détermination automatique des mots-clés. Les éléments de la carte finale seraient alors calculés d'après un système de vecteurs comme le *Reflexive Random Indexing* plutôt que sur un simple comptage d'occurrences [GHALI et al. 12].

3. Expérimentation

Le projet de réalisation du LICI est parti d'une demande pratique des utilisateurs des logiciels de Médialexie. En effet, certains dyslexiques utilisent beaucoup l'outil de création de carte heuristique afin d'organiser leurs idées, de résumer un texte ou leurs leçons. Ces personnes ont exprimés le besoin d'un outil pour le faire automatiquement ou au moins les aider dans cette tâche. La figure 2 est extraite d'une carte heuristique réalisée par un élève dyslexique pour résumer sa leçon d'histoire (on notera la possibilité d'attacher des documents à un élément de la carte, représentés par l'icône du trombone).

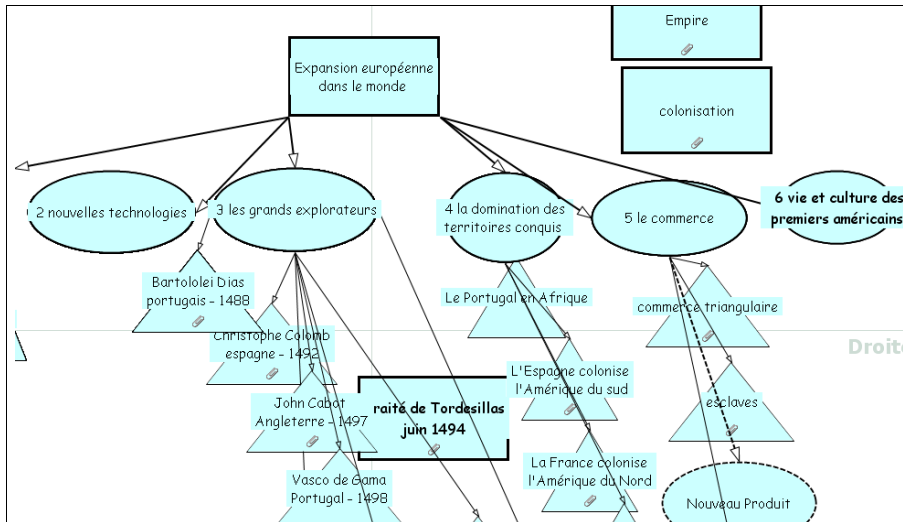


Figure 2. Carte heuristique réalisée par un élève dyslexique

Nous allons lancer une étude pratique auprès des utilisateurs dyslexiques de Médialexie en leur demandant de réaliser manuellement des cartes à partir de textes donnés. Nous proposerons plusieurs types de textes différents : leçon, article, littéraire. Nous procéderons à une analyse qualitative des cartes obtenues afin de déterminer un ou plusieurs types de cartes adaptées au public dyslexique. Nous choisirons ainsi le traitement automatique et la représentation permettant de s'approcher le plus des résultats manuels. Une fois le prototype réalisé, il sera testé par ce même public dyslexique en relation avec l'entreprise Médialexie afin de procéder à des modifications ou des ajustements. La version finale du LICI utilisera la même interface que l'éditeur de carte heuristique déjà existant et permettra aux utilisateurs de travailler directement sur les cartes générées automatiquement.

Remerciements

Les auteurs tiennent à remercier le LRL et l'entreprise Médialexie qui accueillent Mario Laurent ainsi que l'Association Nationale Recherche et Technologie qui a permis ces recherches en accordant une bourse CIFRE pour ce projet de thèse.

3.0. Bibliographie

- [COIRIER et al. 96] Coirier, P., Gaonac'h, D., Passerault, J.-M., *Psycholinguistique textuelle: une approche cognitive de la compréhension et de la production des textes*, Paris, A. Colin, 1996.
- [GHALI et al. 12] Ghali, A. E., Hromada, D., Ghali, K. E., Enrichir et raisonner sur des espaces sémantiques pour l'attribution de mots-clés, *Actes de la conférence TALN 2012*, Grenoble, 2012.
- [HABIB & JOLY-POTTUZ 08] Habib, M., Joly-Pottuz, B., Dyslexie, du diagnostic à la thérapeutique: un état des lieux, *Revue de Neuropsychologie*, vol. 18, n°4, 2008, p. 247-325.
- [HECKER & ENGSTROM 07] Hecker, L., Engstrom, E. U., Assistive Technology and Individuals with Dyslexia, in *Multisensory Teaching of Basic Language Skills (2nd Edition)* (Judith R.), Paul H. Brookes Publishing Company, 2005.
- [INSERM (dir.) 07] INSERM (dir.), *Dyslexie, dysorthographe, dyscalculie: bilan des données scientifiques*, Paris, INSERM, 2007.
- [MARGARIDO et al. 08] Margarido, P. R., Pardo, T., Antinio, G., Fuentes, V., Aires, R., Aluisio, S., Fortes, R. P. M., Automatic Summarization for Text Simplification: Evaluating Text Understanding by Poor Readers, *Actes de la conférence XIV Brazilian symposium on Multimedia and the Web, WebMedia'08*, Vila Velha, Brazil, 2008.

- [NANDHINI & BALASUNDARAM 11] Nandhini, K., Balasundaram, S.R., Improving Readability of Dyslexic Learners through Document Summarization, *Actes de la conférence International Conference on Technology for Education*, IEEE, 2011.
- [TORRES-MORENO 11] Torres-Moreno, J.-M., *Résumé automatique de documents : une approche statistique*, Paris, Hermes, 2011.
- [ZORZI et al. 12] Zorzi, M., Barbiero, C., Facoetti, A., Lonciari, I., Carrozzi, M., Montico, M., Bravar, L., et al., Extra-large letter spacing improves reading in dyslexia, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, n°28, 2012, p.11455-11459.

3.1. Références sur le WEB.

- [JACQUIER-ROUX et al. 10] Jacquier-Roux, M., Lequette, C., Pouget, G., Valdois, S., Zorman, M., B.A.L.E. (Batterie Analytique de Langage Ecrit), *Cogni-Sciences, Ed.*, Grenoble, 2010. (http://www.cognisciences.com/article.php3?id_article=81)
- [MÉDIALEXIE 12] Médialexie, site internet de l'entreprise Médialexie, 2012. (<http://www.medialexie.com/>)
- [TREETAGGER 12] TreeTagger, site internet du projet TreeTagger, 2012. (<http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>)

Texte court, tel qu'il est présent dans les actes de la conférence EIAH 2013 :

Outil d'aide à la compréhension de texte pour des apprenants dyslexiques

Développement d'un outil de cartographie automatique de texte d'aide à la compréhension de texte pour des apprenants dyslexiques

Mario Laurent**, **Thierry Chanier***

* Laboratoire de Recherche sur le Langage,
Université Blaise Pascal, Maison des Sciences de l'Homme
thierry.chanier@univ-bpclermont.fr

** Laboratoire de Recherche sur le Langage et Médialexie
mario.laurent@etudiant.univ-bpclermont.fr

RÉSUMÉ. Les personnes atteintes de troubles du langage, comme la dyslexie, éprouvent de grandes difficultés à lire et à écrire. Ces problèmes persistent lorsqu'il s'agit d'utiliser un environnement bureautique, il est donc nécessaire de leur fournir un ensemble d'outils adaptés. Nous proposons de réaliser un outil, le LICI, permettant d'obtenir automatiquement une carte, conceptuelle ou heuristique, à partir d'un texte. Cet outil facilitera la compréhension de texte en un temps réduit et aidera grandement les dyslexiques lors d'activités de documentation ou d'apprentissage.

MOTS-CLÉS : TAL ; dyslexie ; résumé automatique ; remédiation ; environnement d'apprentissage personnalisé

4. Contexte : dyslexie et assistances technologiques

Les troubles spécifiques du langage et en particulier la dyslexie développementale, constituent une difficulté majeure tout au long de la scolarité des personnes concernées. De

5 à 8% de la population d'une classe d'âge est atteinte et chaque cas est différent [INSERM (dir.) 07]. Les dyslexiques sont fréquemment atteints d'autres troubles, notamment la dysorthographe et la dyscalculie [HABIB & JOLY-POTTUZ 08]. Ces personnes sont en réelle difficulté pour lire et produire des documents écrits. Afin d'assurer leur inclusion sociale, et de favoriser leur cursus scolaire, il est essentiel de mettre à leur disposition des outils d'aide et de remédiation. Depuis le début des années 2000, de nombreux outils ont été adaptés dans ce but : synthèse et reconnaissance vocales, correcteur automatique, prédicteur de mots, et enfin, concepteur de cartes heuristiques. Leur efficacité a été confirmée par plusieurs études [HECKER & ENGSTROM 07]. L'entreprise Médialexie, à l'origine de nos travaux, propose déjà une barre d'outils rassemblant l'ensemble de ces fonctionnalités. Par ailleurs, [NANDHINI & BALASUNDARAM 11] ont montré que les résumés automatiques apportent une aide à la compréhension de texte. Nous savons également qu'une lecture fluide est indispensable pour réaliser l'intégration sémantique d'un texte [COIRIER et al. 96]. Nous espérons compenser la difficulté d'intégration sémantique des dyslexiques en leur proposant directement une représentation graphique des concepts et relations du texte.

5. Cartographie automatique : le LICI

Certains dyslexiques utilisent beaucoup l'outil de création de cartes de Médialexie afin d'organiser leurs idées ou de résumer une leçon. Ce sont eux qui ont exprimé le besoin d'un outil de construction automatique de cartes. L'outil que nous développons, appelé LICI (*Logiciel Intelligent de Cartographie Inversée*), s'adresse d'abord à un public de collégiens. Les cartes générées présenteront les éléments importants pour accéder au sens d'un texte. Les relations pourront être d'ordre sémantique, spatial ou chronologique. L'interface sera simple, les éléments pourront être affichés un par un et lus par une voix de synthèse. Chaque utilisateur aura un profil dynamique et pourra sauvegarder ses préférences : nombre d'éléments affichés, couleurs, espacement des caractères... En effet, d'après [ZORZI et al. 12], une simple augmentation de l'espace entre les caractères améliore de façon significative les performances des dyslexiques. Enfin, à partir de chaque élément de la carte, on pourra accéder à un aperçu du texte présentant ses occurrences surlignées. Au niveau technique, nous utiliserons des méthodes de résumé automatique [TORRES-MORENO 11]. Nous effectuerons un prétraitement : segmentation et étiquetage des graphies, puis tri selon les catégories grammaticales et comptage de fréquences par lemme. Nous tiendrons également compte de la position des éléments dans le texte. Pour déterminer les caractéristiques des cartes adaptées, nous allons lancer une étude pratique auprès d'un public dyslexique qui devra réaliser des cartes manuellement à partir de textes donnés. Une fois le prototype réalisé, il sera testé par ce même public afin de procéder à des ajustements. L'interface du LICI sera similaire à l'éditeur de carte de Médialexie et permettra aux utilisateurs de travailler directement sur les cartes générées.

6. Bibliographie

[COIRIER et al. 96] Coirier, P., Gaonac'h, D., Passerault, J.-M., *Psycholinguistique textuelle: une approche cognitive de la compréhension et de la production des textes*, Paris, A. Colin, 1996.

[HABIB & JOLY-POTTUZ 08] Habib, M., Joly-Pottuz, B., Dyslexie, du diagnostic à la thérapeutique: un état des lieux, *Revue de Neuropsychologie*, vol. 18, n°4, 2008, p. 247-325.

- [HECKER & ENGSTROM 07] Hecker, L., Engstrom, E. U., Assistive Technology and Individuals with Dyslexia, in *Multisensory Teaching of Basic Language Skills (2nd Edition)* (Judith R.), Paul H. Brookes Publishing Company, 2005.
- [INSERM (dir.) 07] INSERM (dir.), *Dyslexie, dysorthographe, dyscalculie: bilan des données scientifiques*, Paris, INSERM, 2007.
- [NANDHINI & BALASUNDARAM 11] Nandhini, K., Balasundaram, S.R., Improving Readability of Dyslexic Learners through Document Summarization, *Actes de la conference International Conference on Technology for Education*, IEEE, 2011.
- [TORRES-MORENO 11] Torres-Moreno, J.-M., *Résumé automatique de documents : une approche statistique*, Paris, Hermes, 2011.
- [ZORZI et al. 12] Zorzi, M., Barbiero, C., Facoetti, A., Lonciari, I., Carrozzi, M., Montico, M., Bravar, L., et al., Extra-large letter spacing improves reading in dyslexia, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, n°28, 2012, p.11455-11459.
-