



## NRC Publications Archive Archives des publications du CNRC

### **Semantic Maps to Aid in Reading** St-Jacques, Claude; Barrière, Caroline

#### **NRC Publications Record / Notice d'Archives des publications de CNRC:**

<https://nrc-publications.canada.ca/eng/view/object/?id=4aa020d6-8181-4dee-b9c3-139363e2d92b>

<https://publications-cnrc.canada.ca/fra/voir/objet/?id=4aa020d6-8181-4dee-b9c3-139363e2d92b>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at

<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site

<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

**Questions?** Contact the NRC Publications Archive team at [PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca](mailto:PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca). If you wish to email the authors directly, please see the first page of the publication for their contact information.

**Vous avez des questions?** Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à [PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca](mailto:PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca).





National Research  
Council Canada

Conseil national  
de recherches Canada

Institute for  
Information Technology

Institut de technologie  
de l'information

# **NRC - CNRC**

---

## *Les cartes sémantiques pour l'aide à la lecture \**

St-Jacques, C., Barrière, C.  
2006

\* publié dans Technologies langagières et apprentissage des langues.  
2006. pp. 67-84. NRC 48763.

Copyright 2006 by  
National Research Council of Canada

Permission is granted to quote short excerpts and to reproduce figures and tables  
from this report, provided that the source of such material is fully acknowledged.

Les cartes sémantiques pour l'aide à la lecture

Claude St-Jacques

Université d'Ottawa

Caroline Barrière

GTLI - CNRC

## Résumé

Les cartes sémantiques permettent de situer un mot inconnu dans un champ sémantique particulier et ainsi aider à sa compréhension. Cet article présente une approche inspirée de la logique floue permettant de faire la construction automatique de cartes sémantiques à partir de l'information contenue dans les définitions d'un dictionnaire. Cette approche est codée dans un module logiciel qui accède au DAFLES (Dictionnaire de l'Apprenant Langue Étrangère et Seconde) et qui s'intégrera à DIDALECT, un didacticiel du français langue seconde. Ce module logiciel inclut aussi une méthode de visualisation des cartes sémantiques permettant à un usager, apprenant du français en situation de lecture, de naviguer au travers des mots sémantiquement reliés et ainsi mieux saisir le sens de ces mots.

## **Les cartes sémantiques pour l'aide à la lecture**

### Un contexte d'apprentissage des langues (DIDALECT)

La présente recherche s'inscrit dans le prolongement d'un projet de didacticiel de lecture (DIDALECT) (Duquette, Desrochers et Szpakowicz 2004) pour l'apprentissage du français langue seconde (L2). L'objectif du projet étant de permettre à un apprenant d'améliorer sa connaissance du français par la lecture de textes, la conception de ce didacticiel intelligent est basée sur deux grandes préoccupations didactiques. D'abord le système doit pouvoir ajuster le niveau de difficulté des textes et des questions de compréhension au stade d'apprentissage atteint par le lecteur. Puis, le logiciel doit fournir des appuis à la lecture suivant les besoins d'enrichissement lexical particuliers à l'utilisateur en ligne.

En soumettant d'entrée de jeu l'apprenant du français L2 à un test d'évaluation, DIDALECT est conçu de manière à établir dès le départ le niveau de compétences linguistiques d'un usager. Le didacticiel peut ensuite présenter à celui-ci un texte calibré selon le degré d'avancement des apprentissages démontré lors de cette évaluation. Bien entendu, les textes du didacticiel ont été préalablement jaugés et classifiés suivant une méthode de calibrage définie par des chercheurs de l'équipe. Vient ensuite la conception d'outils d'aide à la compréhension des textes devant servir de support à l'apprenant dans son processus d'apprentissage langagier. C'est précisément dans le contexte de l'élaboration de ces outils d'appui à la lecture que s'inscrit notre recherche sur les cartes sémantiques.

### Difficultés lexicales en cours de lecture

La lecture d'un texte fait éventuellement surgir chez un apprenant des difficultés de différents ordres : lexicales, morphologiques, syntaxiques et sémantiques. Chacune d'elles nécessitent une intervention didactique adaptée à sa nature. Dans la présente recherche, nous explorons certaines méthodes computationnelles pouvant convenir aux besoins d'un usager face

à des problèmes de compréhension lexicale ou sémantique d'un texte. En contexte d'apprentissage trois choix s'offrent à l'apprenant lorsqu'il veut combler l'absence d'un mot de son vocabulaire : soit il s'enquière du sens du lexème auprès d'une personne pouvant l'en informer (autres étudiants ou professeur en contexte de classe), soit il consulte un dictionnaire, ou soit il tente d'inférer le sens du mot par rapport à sa compréhension du contexte. Bien entendu, en situation d'auto-apprentissage (tel que prévu pour DIDALECT), seule une aide suivant les deux dernières stratégies est envisageable. La lecture d'une définition dans un dictionnaire et le processus d'inférence en contexte sont en fait deux stratégies complémentaires car une ou l'autre utilisée indépendamment est souvent insuffisante.

C'est dans cette optique que s'insère nos travaux, en proposant l'utilisation d'un dictionnaire électronique, sans toutefois restreindre son accès à une définition, mais plutôt en augmentant l'information que l'utilisateur peut accéder dans ce dictionnaire. A plus long terme (travaux futurs), nous explorerons aussi l'impact du contexte de la phrase d'occurrence des mots pour aider le processus d'inférence.

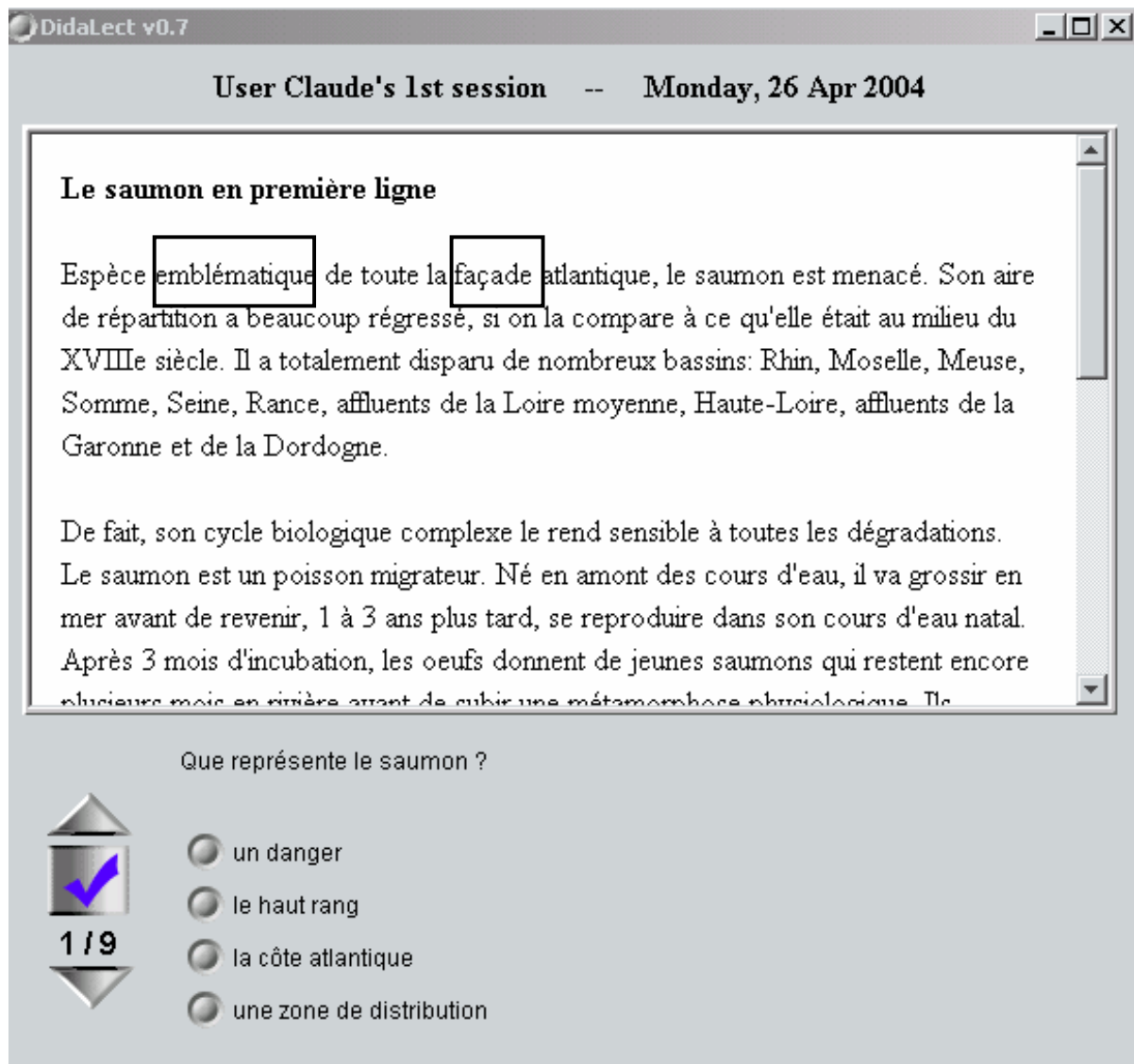
Pour bien saisir le genre de difficultés lexicales auxquelles peut être confronté un apprenant, considérons un texte présenté dans DIDALECT. À la Figure 1, nous encadrons deux mots au tout début du texte : « emblématique » et « façade ». Le premier mot « emblématique » n'a qu'une seule définition dans *Le Trésor de La Langue Française informatisé*<sup>1</sup>(TLFi) : « Qui présente un emblème, qui a trait à un emblème. *Figure, décoration, peinture, dessin, emblématique ; allégorique et emblématique* ». La compréhension du mot est cependant difficile, parce que sa définition génère déjà par la complexité de son vocabulaire une boucle de recherche

---

<sup>1</sup> Voir sur Internet: <http://atilf.inalf.fr/tlfv3.htm>

pouvant entraîner l'utilisateur du dictionnaire dans une pérégrination incessante. En effet, celui-ci doit d'abord saisir dans la définition du TLFi le sens du substantif « emblème » et de l'épithète de substitution synonymique « allégorique ».

Figure 1. Exemple de mots difficiles dans DIDALECT.



The screenshot shows a window titled "DidaLect v0.7" with a subtitle "User Claude's 1st session -- Monday, 26 Apr 2004". The main text area contains the following passage:

**Le saumon en première ligne**

Espèce **emblématique** de toute la **façade** atlantique, le saumon est menacé. Son aire de répartition a beaucoup régressé, si on la compare à ce qu'elle était au milieu du XVIII<sup>e</sup> siècle. Il a totalement disparu de nombreux bassins: Rhin, Moselle, Meuse, Somme, Seine, Rance, affluents de la Loire moyenne, Haute-Loire, affluents de la Garonne et de la Dordogne.

De fait, son cycle biologique complexe le rend sensible à toutes les dégradations. Le saumon est un poisson migrateur. Né en amont des cours d'eau, il va grossir en mer avant de revenir, 1 à 3 ans plus tard, se reproduire dans son cours d'eau natal. Après 3 mois d'incubation, les oeufs donnent de jeunes saumons qui restent encore plusieurs mois en rivière avant de subir une métamorphose physiologique. Ils

Below the text is a question: "Que représente le saumon ?". To the left of the question is a navigation icon with a blue checkmark and the number "1 / 9". To the right are four radio button options:

- un danger
- le haut rang
- la côte atlantique
- une zone de distribution

En fouillant la définition d'« emblème », l'apprenant peut se butter à la difficulté de compréhension de certains mots rares comme « sceptre », « couronne », ou à des expressions très particulières comme « emblème maçonnique », « emblème mythologique ». La même

multiplication des recherches peut se produire pour la compréhension du sens d'« allégorique » qui réfère d'abord à la compréhension du substantif « allégorie ». Bref, il est parfois laborieux pour l'utilisateur d'un dictionnaire d'en arriver, de par l'utilisation traditionnelle de ce type d'outil de référence, à une vue synthétique du champ sémantique de certains mots difficiles comme « emblématique ».

Examinons maintenant le deuxième mot encadré dans notre exemple de la Figure 1.

Selon le TLFi « façade » a plusieurs sens:

1. « Mur extérieur d'un bâtiment. *La façade du côté de la cour. La façade qui regarde la rivière* ».
2. « Partie visible formant un mur. *Une sorte de niche dans la façade à pic du rocher* ».
3. « Au fig. Apparence souvent trompeuse. *La médiocrité intolérable du petit employé, aussi désastreuse que la misère noire de l'ouvrier, la façade fausse* ».

Ce deuxième mot est ipso facto ambiguë. D'abord la polysémie (2 sens propres et 1 figuré) du mot « façade » présente déjà une première difficulté pour l'apprenant. Il devra pouvoir distinguer quel sens est le plus approprié dans son contexte de lecture. En fait, ici, la difficulté est encore plus grande parce que le sens utilisé dans le texte ne fait pas parti des 3 sens montrés. Cela demande un effort supplémentaire d'inférence chez l'apprenant que de trouver un « pont sémantique » entre un des sens donnés et celui du texte.

Même si nous l'illustrons ici, nous n'aborderons pas plus loin cette question d'inférence d'extension de sens et nous nous limiterons à l'étude de la polysémie en supposant que le dictionnaire contient l'information désirée. Même cette polysémie « pure », en supposant qu'elle couvre tous les sens possibles, est un travers à la compréhension d'une langue qui est malaisé à



un apprenant de surmonter, compte tenu des limitations lexicographiques des dictionnaires et de ses modes traditionnels d'utilisation dont il est culturellement difficile de se départir. Cette dernière remarque à propos du mode usuel d'utilisation des dictionnaires laisse entrevoir la nature exploratoire de notre recherche que nous dépeignons dans les sections suivantes comme étant sous certains aspects en rupture avec cette tradition.

#### Aide à la compréhension des mots : les cartes sémantiques

Sans rejeter complètement l'utilisation traditionnelle du dictionnaire, il faut cependant rappeler (Bogaards, 1996, 277-320) que certaines études menées auprès d'apprenants de langue seconde ont montré que l'utilisation du dictionnaire est trop souvent difficile, voir même déroutante, et peut nuire en contexte d'écriture. Une première étude de Hulstijn (2000) tend à démontrer que la consultation d'un dictionnaire en situation d'écriture est plus propice à la mémorisation d'un nouveau mot que cela ne peut l'être en situation de lecture. L'exercice d'écriture force un individu à faire appel à plusieurs processus mentaux de compréhension d'une langue ce qui faciliterait dans ce contexte la rétention du nouveau mot élucidé par la consultation d'un dictionnaire. Par ailleurs, d'autres chercheurs (Nagy, Herman, Anderson, 1985, 233-253) soutiennent que l'apprentissage de nouveaux mots est plus fréquemment l'incidence d'une lecture ou d'une écoute répétitive. Le résultat de plusieurs études empiriques (Chun, Plass, 1996) (Hulstijn, Hollander, Greidanus, 1996) (Zimmerman, 1997, 121-140) confirment cette dernière thèse.

Outre l'exercice répété d'un nouveau mot, les activités permettant d'augmenter la connaissance du vocabulaire d'une langue doivent aussi prévoir le développement d'une fluidité dans l'usage. Néanmoins la fluidité dépend du vocabulaire déjà vu (Hunt, Beglar, 1998) au fil d'exercices répétitifs et l'utilisation de petites cartes mobiles (à l'écran ou sur un morceau de

papier), montrant le groupe auquel appartient le mot recherché, accentue la vitesse d'exécution de l'apprenant dans le cadre de la compréhension d'un extrait de texte.

Cette idée de groupement est à la source des cartes sémantiques. L'utilisation des cartes sémantiques comme stratégie de pré-lecture pour activer des connaissances préalables à la compréhension d'un texte n'est pas une technique nouvelle en soi. Les premières études sur cette technique d'enseignement, aidant à combler les vides laissés par des mots inconnus en situation de lecture, datent des années 70 et 80 (Johnson, Pittelman, Heimlich, 1986, 778-782 ; Antonacci, 1991, 174-194). Les cartes sémantiques combinent plusieurs méthodologies en usage dans le domaine de l'apprentissage des langues : les listes de mots, leur regroupement, leur association par concept, l'imagerie visuelle, l'utilisation des dictionnaires, etc (Svenconis, 1994). Quoique l'utilisation des cartes sémantiques (Svenconis, 1994) ou des techniques assistées par ordinateur pour l'élucidation de mots inconnus (Hulstijn, 2000, 32-43) en apprentissage des L2 demeurent des approches encore très peu étudiées ou même explorées par les spécialistes de cette discipline, nous nous appuyons sur certains travaux en faveur de l'utilisation de ces cartes.

Certains chercheurs (Sinatra, Stahl-Gemake, Berg, 1984, 22-29 ; Antonacci, 1991) ont démontré qu'une représentation visuelle avec des cartes sémantiques aide un mauvais lecteur à réorganiser de nouvelles informations autour du réseau conceptuel qu'il connaît déjà. Plusieurs études empiriques (Crow and Quigley, 1985, 497-513) (Brown and Perry, 1991, 655-670) ont démontré que le processus de préhension de la sémantique d'un nouveau mot est essentiel à une stratégie d'apprentissage du vocabulaire d'une L2.

Bref, seule la permissivité de la technique plaide en défaveur des cartes sémantiques. En effet, cette méthode de structuration laisse libre cours à la navigation dans une carte et un tel

environnement peut produire une surcharge cognitive chez l'apprenant si cette ressource est mal utilisée (Svenconis, 1994).

### Automatisation de la construction des cartes sémantiques

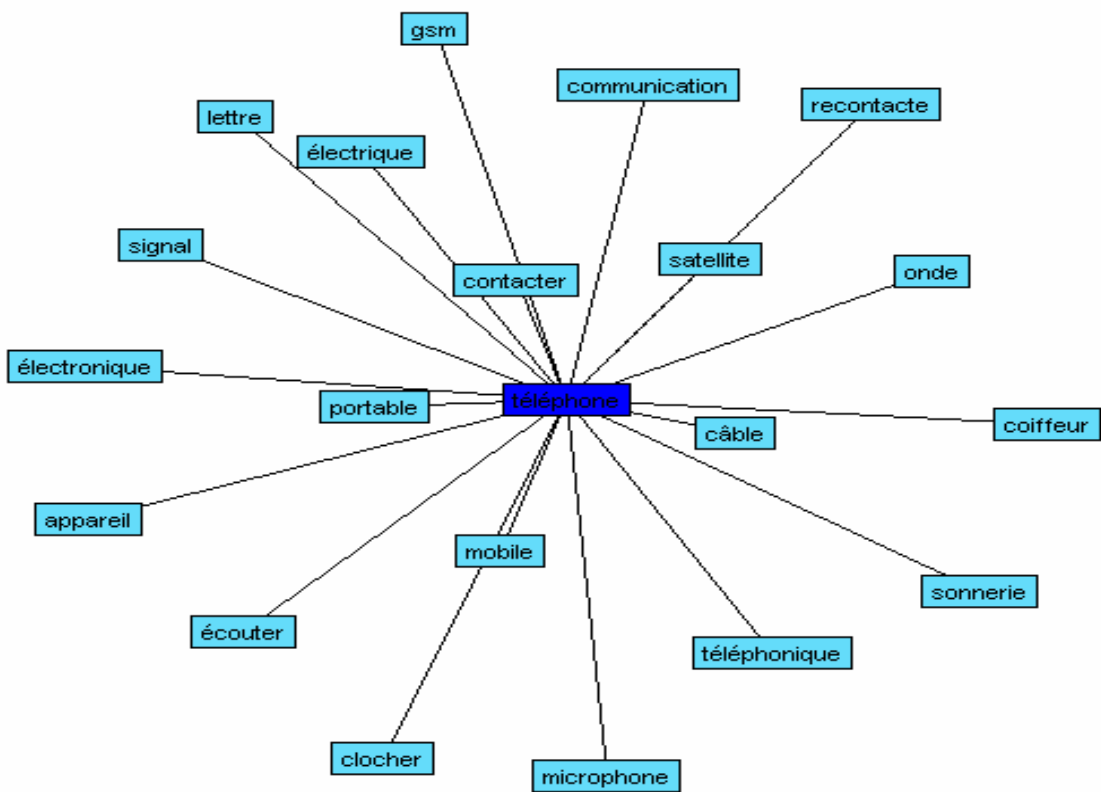
Toutes les études précédemment citées font l'utilisation de cartes sémantiques manuellement construites. Ce processus est long et laborieux, et la contribution de notre présent travail est de proposer une méthode de construction automatique de cartes sémantiques. Nous focusons donc sur l'aspect construction dans cette section et discuterons de l'utilisation et navigation dans la prochaine section.

Dans le cadre de l'élaboration automatique de cartes sémantiques dans un contexte d'apprentissage des langues, il est important que la ressource que nous utilisons réponde à deux critères. D'abord que cette ressource ait été pensée en contexte d'apprentissage, ensuite qu'elle contienne l'ensemble des mots qui pourraient poser problème à un apprenant. Nous soulignons donc qu'une obligeante collaboration avec les auteurs du DAFLES<sup>1</sup> (Dictionnaire de l'Apprenant du Français Langue Étrangère et Seconde) nous a permis de profiter de cette ressource dans nos travaux exploratoires. Le DAFLES est conçu pour des apprenants et quoiqu'il ne contient présentement que les mots les plus fréquents de la langue, une partie de l'équipe de DIDALECT contribut à l'ajout des mots difficiles contenus dans les textes choisis.

Prenons un exemple tiré de l'un des textes de DIDALECT, soit la phrase suivante :  
« Voilà qu'une étude empirique le confirme : si vous utilisez votre **téléphone**<sup>2</sup> mobile tout en conduisant, vous augmentez de plus du tiers (38 p. 100) vos risques d'accidents de la route ». Il est évident que la saisie du mot « téléphone » est essentielle à la compréhension de cette phrase. Nous donnons à la Figure 2 la carte sémantique des associations lexicales du mot « téléphone »

construite automatiquement par notre logiciel à partir des données du DAFLES, en suivant l'algorithme suivant :

Figure 2. Associations lexicales autour de « téléphone ».



1. Un prétraitement nous permet d'abord de limiter le vocabulaire aux termes signifiants du dictionnaire (filtrage des mots outils et adverbes) ;
2. Pour chaque terme retenu, nous procédons au comptage de ses fréquences d'apparition dans une entrée lexicale du dictionnaire afin d'établir son degré d'appartenance à l'ensemble flou représentatif de cette entrée ;
3. Un pseudo-thésaurus flou, que nous appellerons aussi réseau d'associations lexicales, est généré automatiquement à partir du DAFLES selon un calcul basé sur le poids de co-occurrence de deux termes dans une entrée.

Prenons un exemple pour bien illustrer les deux premières étapes. Supposons le mot « culture » consulté par l'utilisateur de DIDALECT, et regardons les trois premiers sens du DAFLES:

- i. « la culture est tout ce qui se rapporte aux activités intellectuelles de divertissement, de réflexion : les arts (le théâtre, la musique, la littérature, le cinéma, la peinture, etc.), la philosophie, etc. » ;
- ii. « la culture est l'ensemble des connaissances d'une personne dans le domaine des arts (le théâtre, la musique, la littérature, le cinéma, la peinture, etc.), de la philosophie, etc. » ;
- iii. « une culture est un ensemble d'habitude propre à un peuple, à une organisation ».

La première étape de notre algorithme utilise des traitements classiques en science de l'information (Salton & Lesk, 1971, 115-141) pour la mise en forme des données tel le filtrage des mots outils et la réduction des formes canoniques d'un même mot au féminin, au pluriel ou bien conjugué. Ainsi, les mots tels « de, la, un, est » seront filtrés. L'étape suivante compte les apparitions d'un mot dans une définition en prenant chacune d'elle comme une entrée lexicale séparée.

Table 1 - Comptage des occurrences d'un terme dans une entrée du DAFLES

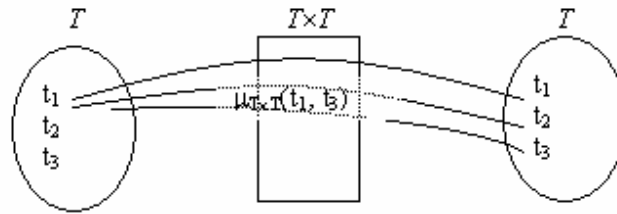
<i>E</i>	<i>t<sub>j</sub></i>
Culture <sub>1</sub>	culture (2), rapporte (1), activité (1), intellectuelle (1), divertissement (1), réflexion (1), art (1), théâtre (1), musique (1), littérature (1), cinéma (1), peinture (1), philosophie (1)
Culture <sub>2</sub>	culture (2), ensemble (1), connaissance (1), personne (1), domaine (1), art (1), théâtre (1), musique (1), littérature (1), cinéma (1), peinture (1), philosophie (1)
Culture <sub>3</sub>	culture (2), ensemble (1), habitude (1), propre (1), peuple (1), organisation (1)

À partir des données fournies par le comptage des occurrences d'un terme dans une entrée tel que montrées à la Table 1, nous pouvons ensuite exécuter la troisième étape de notre algorithme pour produire automatiquement un pseudo-thésaurus flou à partir du DAFLES. L'idée derrière cette génération automatique d'un thésaurus d'associations floues, entre les termes d'un dictionnaire pris deux à deux, nous vient de l'application de Miyamoto (1990) de la théorie des ensembles flous en science de l'information.

Suivant cette thèse, les concepts que nous formons en notre esprit à propos des objets du monde, ne peuvent nous être accessibles simplement par l'intermédiaire d'un artefact. Malgré cela, l'être humain est toujours capable d'exprimer des concepts en langage naturel. L'hypothèse derrière cet algorithme est de dire que si une machine ne peut accéder directement à nos concepts, elle peut, si nous la dotons d'une stratégie qui s'apparente à la nôtre, les identifier indirectement à l'aide de leurs expressions dans une langue naturelle. D'où l'idée (Miyamoto, 1990) qu'un concept est exprimé par une association floue de termes. Ceci signifie que pour évoquer un concept nous utilisons un certain nombre de mots que nous associons ensemble avec certaines variations d'une fois à l'autre dans le regroupement dépendamment du contexte.

Pour revenir à notre application, admettons en suivant cette ligne de pensée qu'un concept est exprimé dans le DAFLES par une association entre certains termes de son vocabulaire. Suivant cette perspective d'agrégation variable de mots en un concept, si nous appelons  $T$  l'ensemble des termes de ce vocabulaire, nous traduisons formellement cette association floue entre des termes comme le produit cartésien  $T \times T$ .

Figure 3. L'ensemble flou  $T \times T$  généré par l'association flexible entre des termes du DAFLES.



Le processus d'association floue entre les termes est représenté à la Figure 3 par la mise en correspondance de l'ensemble du vocabulaire  $T$  avec lui-même. Prenons un exemple avec le terme  $t_1$ , nous pouvons déterminer les termes avec lesquels il coexiste dans une entrée lexicale, c'est-à-dire selon cette figure les termes  $t_1$ ,  $t_2$  et  $t_3$ . Sachant maintenant que le vocabulaire  $T$  du DAFLES est un ensemble déterminé et que, suivant la théorie de la logique floue (Bouchon-Meunier, 1994) le produit cartésien entre deux ensembles ordinaires est un mode éventuel de génération d'ensembles flous, alors nous pouvons caractériser cette relation floue entre deux termes par son degré d'appartenance  $\mu_{T \times T} \rightarrow [0,1]$  à l'ensemble flou  $T \times T$ <sup>2</sup>.

La cooccurrence de deux termes peut donner lieu à plus d'une relation floue d'association. Le calcul de génération automatique du pseudo-thésaurus flou est basé sur la relation de similarité floue (Miyamoto, 1990) :

$$R(t_i, t_j) = \frac{\sum_k^M \min(t_{ik}, t_{jk})}{\sum_k^M \max(t_{ij}, t_{jk})}$$

---

<sup>2</sup> C'est la particularité des ensembles flous que d'avoir des objets qui leur appartiennent selon un degré d'appartenance défini sur un intervalle de valeurs  $[0,1]$  alors qu'un objet appartient ou pas à un ensemble ordinaire (soit une valeur de 1 ou 0).

Cette formule nous dit que pour chaque association de termes par couple  $(t_i, t_j)$ , le poids de la relation floue de similarité  $R(t_i, t_j)$  nous est donné par la somme des poids de cooccurrence minimaux  $\sum_k^M \min(t_{ik}, t_{jk})$  parmi toute les entrées  $k$  de la matrice des entrées lexicales du dictionnaire  $M$  divisée par la somme des poids de cooccurrence maximaux  $\sum_k^M \max(t_{ij}, t_{jk})$  parmi les mêmes entrées  $k$  de la matrice  $M$ . Si nous revenons à la Table 1 et prenons « culture » et « philosophie » comme termes  $t_i$  et  $t_j$ , les trois entrées  $E$  de ce tableau comme matrice  $M$ ,

alors nous obtenons la relation de similarité floue  $R(t_i, t_j) = \frac{\sum_{k=1}^3 \min(2, 1), \min(2, 1)}{\sum_{k=1}^3 \max(2, 1), \max(2, 1)}$ . Il en résulte

$R(t_i, t_j) = 2/4 = 0.5$ , c'est-à-dire un degré d'appartenance de  $\mu_{R(t_i, t_j)} = 0.5$  à l'ensemble flou  $R(t_i, t_j)$  généré par la relation floue de similarité entre les termes pris deux à deux dans le DAFLES.

Le calcul de la relation de similarité floue entre tous les mots du vocabulaire du DAFLES produit un pseudo-thésaurus flou, dont les poids forment une matrice  $R$ . C'est une portion de cette matrice qui contient les associations lexicales que nous pourrions présenter sous forme de carte sémantique dans laquelle il est possible aussi de naviguer.

### Navigation dans les cartes sémantiques en contexte de lecture

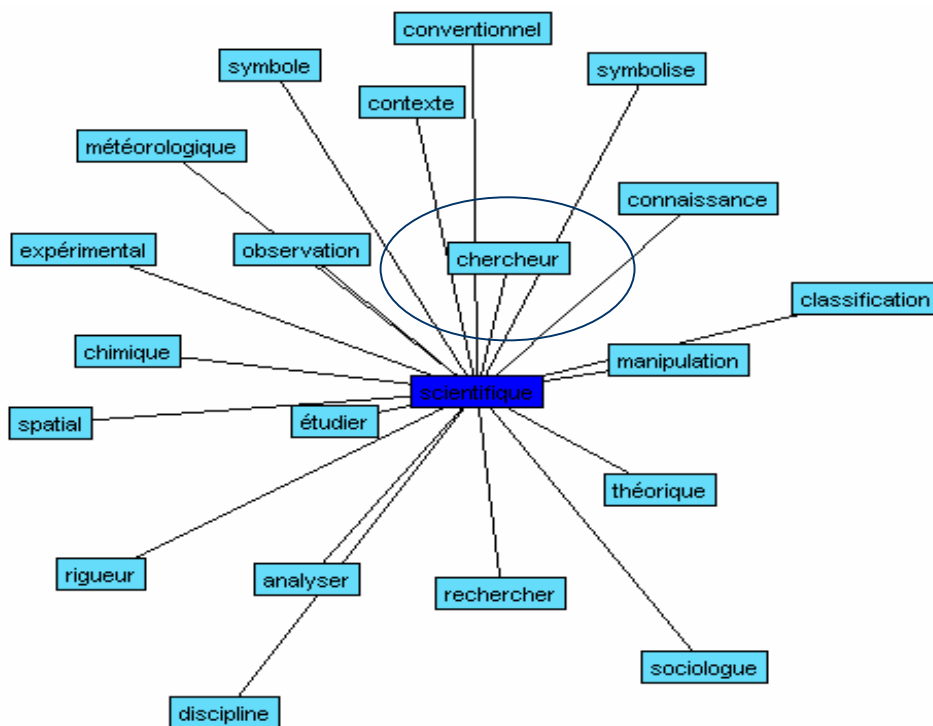
Si les études empiriques tendent à démontrer que l'utilisation des cartes sémantiques facilite l'apprentissage de nouveaux mots (Crow, Quigley, 1985 ; Brown, Perry, 1991), il nous faut ajouter à notre exploration informatique l'objectif d'accorder au système une certaine souplesse en considération de la fluidité à acquérir par l'utilisateur. Comme une lacune lexicale est souvent difficile à combler par la consultation d'une seule entrée d'un dictionnaire, nous



voulons permettre à l'apprenant de naviguer dans la ressource lexicale en trouvant plusieurs mots associés au lexème recherché, afin de le rapprocher le plus possible du contexte sémantique de sa lecture.

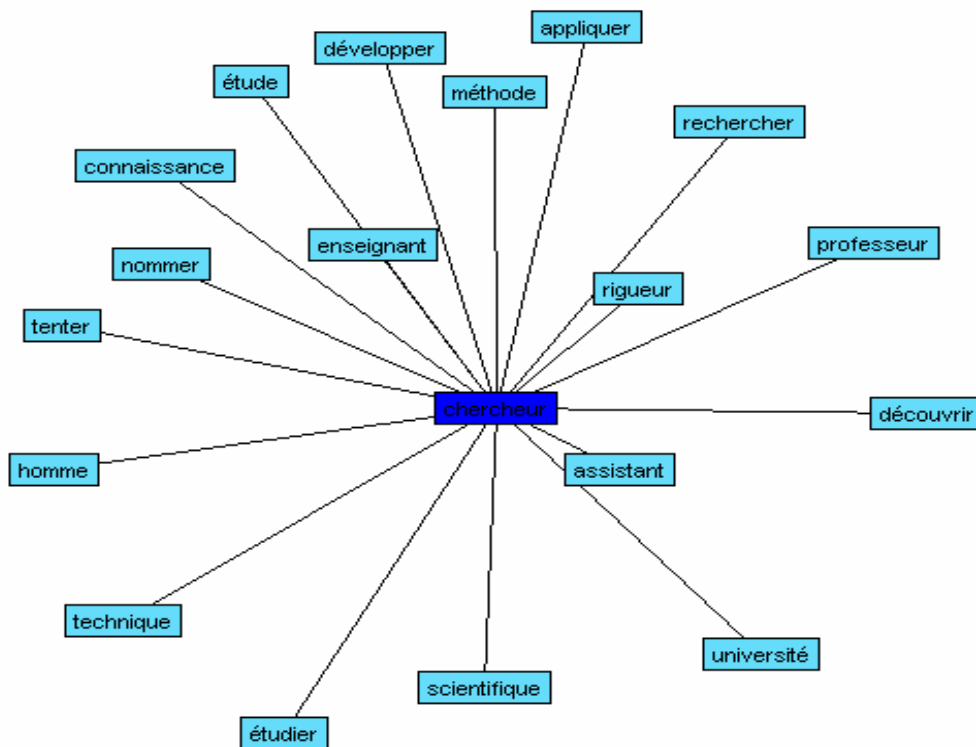
Ainsi, du pseudo-thésaurus flou généré automatiquement à partir du DAFLES par notre programme (matrice R présentée à la section précédente), nous pouvons ensuite produire une carte sémantique des associations lexicales pour un mot consulté par l'utilisateur de DIDALECT. La Figure 4 montre la carte sémantique du mot « scientifique » à partir de ce pseudo-thésaurus du DAFLES. Appelons le mot choisi par l'utilisateur, le « mot-requête ». Il faut noter que la visualisation du graphe est conçue de telle manière que la distance entre un mot et le mot-requête est inversement proportionnelle à leur poids d'association. Ce principe de visualisation est simple et intuitif. Ainsi, suivant le calcul de similarité floue, le mot « étudier » a un poids d'association beaucoup plus grand avec « scientifique » que ne l'a le mot « discipline » situé à une distance beaucoup plus éloignée.

Figure 4. Carte sémantique du mot-requête « scientifique ».



Il est possible pour un utilisateur de naviguer dans une carte sémantique. En cliquant par exemple sur le mot « chercheur » dans la carte sémantique de la Figure 4, une nouvelle carte des associations lexicales est produite à partir du pseudo-thésaurus avec redéploiement autour du mot-requête « chercheur » tel que montré à la Figure 5. Cette particularité de notre système permet de répondre à la deuxième exigence que nous nous sommes fixées soit une souplesse informatique dans l'exploration des cartes sémantiques permettant à l'utilisateur d'acquérir une certaine fluidité dans l'usage d'un nouveau mot.

Figure 5. Redéploiement d'une carte autour du mot-requête « chercheur ».



#### A la découverte de mots nouveaux sans s'égarer

Quoique les réseaux d'associations lexicales soient intéressants en soi, ils peuvent aussi aider à la construction d'un deuxième type de cartes sémantiques qui cette fois nous plonge

directement dans la problématique de la polysémie. En prenant pour acceptation que la signification d'un mot nous est donnée par l'usage (Wittgenstein, 1968 : 43), nous considérons que la compréhension sémantique vient de la désambiguïsation de la polysémie d'un terme ou simplement de l'inférence de son sens unique par la convergence des usages. En effet, il sera intéressant pour un apprenant du français de voir comment un certain mot est au cœur d'un ensemble de définitions. Ce réseau de définitions pertinentes aide à la désambiguïsation en proposant uniquement les sens des définitions associés à la requête.

Notre stratégie est encore une fois fondée sur l'exploitation du dictionnaire comme d'un corpus textuel. Nous avons démontré théoriquement ailleurs (St-Jacques & Barrière 2004) que l'inférence dictionnaire utilisée sur un dictionnaire électronique permet de briser les limites imposées par la structure de ses entrées, dans le but d'exploiter l'ensemble de ses définitions et de ses exemples en contexte pour établir quelles entrées sont les plus pertinentes à une recherche d'information.

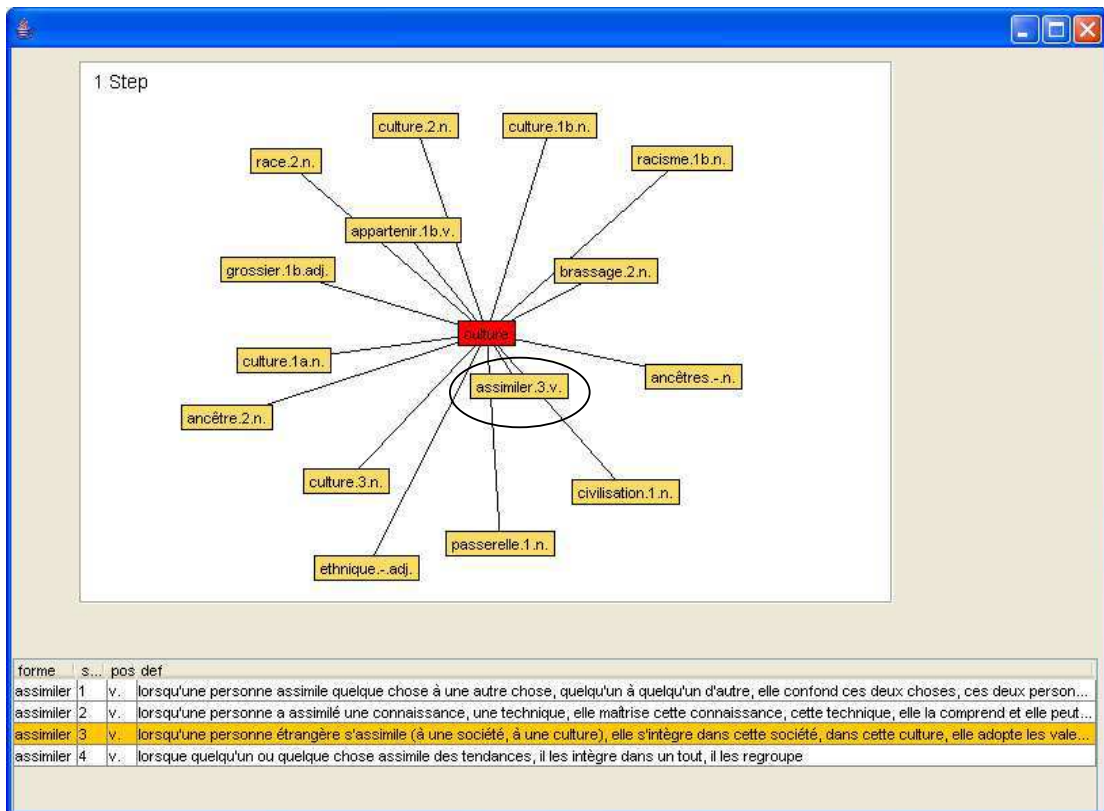
Notre méthode débute avec le mot-requête inconnu de l'apprenant dont nous tentons dans un premier temps d'élargir le champs sémantique à l'aide des mots qui lui sont associés dans le pseudo-thésaurus. L'augmentation de la requête sur le pseudo-thésaurus flou s'effectue au moyen de l'opération de composition *max-min* (représentée aussi par le symbole  $\circ$ ) entre deux relations (Klir, Yuan, 1995). Soit la relation  $Q$  entre les termes de la requête de l'utilisateur et ceux du pseudo-thésaurus, soit aussi la relation  $R$  précédemment décrites pour la génération de ce même thésaurus flou, alors la requête augmentée  $S$  nous est donnée par la formule :  $S = Q \circ R^3$ .

---

<sup>3</sup> Voir l'Annexe A pour un exemple de calcul.

Une requête augmentée  $S = (t_1, t_2, t_3, t_4)$  peut ensuite être appliquée, à la matrice  $E$  des vecteurs de termes décrivant chacune des entrées lexicales du DAFLES, encore une fois par l'opération de composition *max–min* de deux relations. Soit la relation  $S$  décrite comme celle de la requête augmentée et la relation  $E$  que nous présentons maintenant comme le poids des termes associés à une entrée alors le résultat  $A$  de cette application nous est donné par  $A = S \circ E^4$ .

Figure 6. Définitions associées au mot « culture ».



<sup>4</sup> Voir l'Annexe B pour un exemple de calcul.

Dans l'exemple de la Figure 6 l'utilisateur pourra voir comment un des sens du mot « assimiler » se définit par sa relation au mot culture. Le programme identifie automatiquement le sens de ce mot polysémique qui se rapporte à la requête sur le dictionnaire. Cette identification du sens pertinent à un contexte facilite l'apprentissage de nouveaux mots par l'utilisateur du didacticiel en évitant un égarement de sa lecture pouvant être engendré par la préhension de définitions inappropriées.

### Conclusion et futurs travaux

Notre expérimentation de la génération automatique de cartes sémantiques à partir du DAFLES nous amène à conclure que les méthodes flexibles d'exploitation d'un dictionnaire électronique, comme ceux rendus disponibles par la logique floue, ouvre une nouvelle ère de développement pour des outils d'aide à la navigation dans un dictionnaire. Non seulement la base de données du dictionnaire peut être utilisée comme un corpus textuel pour des applications comme la génération automatique de cartes sémantiques, mais en plus la structure lexicographique du dictionnaire peut être exploitée sous un nouveau jour. Le regroupement de plusieurs entrées pertinentes pouvant fournir de l'information à propos de l'objet d'une fouille dans le dictionnaire donne une esquisse du potentiel innovateur d'un tel outil. Le champ sémantique d'un mot peut être fourni en débordant de la structure traditionnelle qui se limite à l'entrée définitionnelle et de ses exemplifications d'usage.

Notre expérimentation sur la génération automatique de cartes sémantiques montre qu'il est possible d'intégrer cet outil à un didacticiel comme DIDALECT. Par ailleurs, le modèle générique de nos algorithmes de calcul permet de penser une adaptabilité à n'importe quel dictionnaire. Bien entendu, cette première expérimentation ouvre la porte à plusieurs questions concernant l'exploitation des cartes sémantiques comme appui à la lecture pour l'apprentissage

d'une langue seconde. Il faut examiner par exemple avec les spécialistes de l'enseignement des langues quel nombre d'associations lexicales il est préférable de montrer à l'apprenant pour l'aider dans sa lecture. En incorporant cet outil à DIDALECT, nous pouvons aussi nous questionner sur l'intérêt de souligner pour l'apprenant dans le texte les associations lexicales pareilles à celles générées automatiquement à partir du dictionnaire. L'une des pistes à explorer en ce sens dans de futurs travaux serait de faciliter la planification d'une étude empirique menée par des spécialistes de l'enseignement sur l'utilisation des cartes sémantiques en gardant la trace des actions des utilisateurs. À partir d'un vocabulaire contrôlé du dictionnaire, il serait possible par exemple de contrôler en post-lecture si les cartes sémantiques ont permis à l'utilisateur d'élucider un mot consulté.

D'autres questions demeurent aussi ouvertes concernant l'utilisation de l'outil que nous proposons. Par exemple, quelles seraient les définitions du dictionnaire les plus importantes à retenir lors d'une requête ? Faut-il donner plus d'importance par exemple aux définitions levant le voile sur les faux-amis ? Comment pouvons-nous trouver l'intersection entre l'information provenant du dictionnaire et celle du texte pour faciliter la compréhension du lecteur ? Bref en explorant de nouveaux usages des dictionnaires électroniques, nous devons ensuite poursuivre nos recherches du côté de l'applicabilité de ces outils à la didactique des langues.

## Bibliographie

- Antonacci, P. A. (1991). Students Search for Meaning in the Text Through Semantic Mapping. *Social Education*, 55, 174-194.
- Bogaards, P. (1996). Dictionaries for Learners of English. *International Journal of Lexicography*, 9, 277-320.
- Bouchon-Meunier, B. (1994). *La logique floue*. 2<sup>e</sup> éd., Paris, France : PUF.
- Brown, T. S. & Perry Jr., F. L. (1991). A comparison of three learning strategies for ESL vocabulary acquisition. *TESOL Quarterly*, 25, 655-670.
- Chun, D. & Plass, J. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal*, 80, 183-198.
- Crow, J. T. & Quigley, J. R. (1985). A semantic field approach to passive vocabulary acquisition for reading comprehension. *TESOL Quarterly*, 19, 497-513.
- Duquette, L. & Desrochers, A. & Szpakowicz S. (2004). Didialect ...
- Hulstijn, J. H. (2000). The use of computer technology in experimental studies of second language acquisition: A survey of some techniques and some ongoing studies. *Language Learning & Technology*, Vol. 3, No.2, 32-43.
- Hulstijn, J. H., Hollander, M. & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, 80, 327-339.
- Hunt, A. & Beglar D. (1998). Current Research and Practice in Teaching Vocabulary. *The Language Teacher Online*, 22.01 (January 1998). Online. Available at <http://jalt-publications.org/tlt/files/98/jan/hunt.html>. Accessible le 26/05/04.

- Johnson, D. D., Pittelman, S. D. & Heimlich, J. E. (1986). Semantic mapping. *The Reading Teacher*. 39, 778-782.
- Miyamoto, S. (1990). *Fuzzy Sets in Information Retrieval and Cluster Analysis*. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Nagy, W. E., Herman, P. & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, 20, 233-253.
- St-Jacques, C. & Barrière, C. (à paraître). L'inférence dictionnaire : De la créativité poétique à celle du raisonnement flou. *Cahiers de lexicologie*.
- Salton, G. & Lesk M. E. (1971). Information Analysis and Dictionary Construction. In Salton, G. (Ed.), *The SMART Retrieval System* (pp. 115-141). Englewood Cliffs, NJ: Prentice Hall inc.
- Sinatra, R. C., Stahl-Gemake, J. & Berg, D. N. (1984). Improving reading comprehension of disabled readers through semantic mapping. *The Reading Teacher*. 38, 22-29.
- Svenconis, D. J. (1994). *An investigation into the teaching of second-language vocabulary through semantic mapping in the hypertext environment*. Thèse de doctorat. The Catholic University of America, Washington, DC.
- Wittgenstein, L. (1968). *Philosophical Investigations*. Trans. by G. E. M. Anscombe, Oxford : Basil Blackwell, cité selon l'usage par numéros d'aphorisme.
- Zimmerman, C. B. (1997). Do reading and interactive vocabulary instruction make a difference?: An empirical study. *TESOL Quarterly*, 31, 121-140.



## Annexe A

Prenons l'exemple fictif suivant : soit la requête initiale  $Q(t_1, t_2)=[1, 1]$  et un pseudo-

thésaurus  $R=$  
$$\begin{matrix} & t_1 & t_2 & t_3 & t_4 \\ \begin{matrix} t_1 \\ t_2 \end{matrix} & \begin{bmatrix} 1 & .8 & .02 & .5 \\ 0 & 1 & .4 & 0 \end{bmatrix} \end{matrix}$$
. Ainsi,  $S = Q \circ R = \max [\min(1, 1), (1, 0)], \max$

$$[\min(1, .8), (1, 1)], \max [\min(1, .02), (1, .4)], \max [\min(1, .5), (1, 0)] = \begin{matrix} t_1 & t_2 & t_3 & t_4 \\ [1 & 1 & .4 & .5] \end{matrix}$$

## Annexe B

Poursuivons l'exemple fictif : soit la requête augmentée  $Q(t_1, t_2, t_3, t_4)=[1, 1, .4, .5]$  et la

matrice des entrées  $E=$  
$$\begin{matrix} & t_1 & t_2 & t_3 & t_4 \\ \begin{matrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{matrix} & \begin{bmatrix} 0 & 0.5 & 0 & 0 \\ 0 & 0.33 & 0 & 0 \\ 0.33 & 0.7 & 0 & 0 \\ 0 & 0 & 0.2 & 0 \end{bmatrix} \end{matrix}$$
. Le résultat de l'application de cette

requête augmentée nous est alors donné par  $A = S \circ E = \max [\min(1, 0), (1, .5), (.4, 0), (.5, 0)],$   
 $\max [\min(1, 0), (1, .33), (.4, 0), (.5, 0)], \max [\min(1, .33), (1, .7), (.4, 0), (.5, 0)], \max [\min(1, 0),$

$$(1, 0), (.4, .2), (.5, 0)] = \begin{matrix} e_1 & e_2 & e_3 & e_4 \\ [.5 & .33 & .7 & .2] \end{matrix}$$

---

<sup>1</sup> Nous soulignons la collaboration entre Lise Duquette de l'Université d'Ottawa et l'équipe de Serge Verlinde et Thierry Selva respectivement auteurs et concepteur informatique du DAFLES de l'Université de Leuven, qui nous a permis de bénéficier gracieusement de la base de données lexicales de ces derniers pour notre expérimentation.

<sup>2</sup> Évidemment « téléphone » est un mot facile à comprendre pour un apprenant anglophone du français L2 puisqu'il est pour eux un des « congénères interlinguaux ». Notre choix est orienté vers cet exemple très simple puisque seuls les mots les plus fréquents du français sont pour le moment inclus dans la base de données lexicales du DAFLES.