

UNE MÉTHODE POUR L'ALIGNEMENT DE TERMES COMPLEXES PLURILINGUES DANS LES TEXTES SPÉCIALISÉS

*Jacques Ladouceur, Guylaine Cochrane,
Ligia Maria Caf  de Miranda**

R SUM : Depuis quelques ann es, plusieurs chercheurs se sont int ress s   l'extraction et   l'exploitation des connaissances contenues dans des corpus de traductions. Ils voyaient l , entre autres choses, le moyen de fournir au traducteur des outils pour le rep rage d' quivalents interlinguistiques. L'exploitation de traductions pose cependant un probl me parce que les  quivalents ne sont pas r dig s spontan ment alors que le premier crit re consid r  par le terminologue est la langue dans laquelle a  t   crit le document qu'il explore (langue d'origine ou langue de traduction). Cela nous a amen    nous poser la question suivante: est-il possible de rep rer des  quivalents interlinguistiques   partir de textes originaux de langues diff rentes? La r ponse est oui.   partir de textes r dig s spontan ment en fran ais, en anglais et en portugais, il a  t  possible de trouver rapidement, pour bon nombre de termes complexes d'une langue, les  quivalents dans les deux autres langues. La seule condition   remplir c'est d'utiliser des textes qui appartiennent   un m me domaine et qui ont la m me fonction de communication. Ce rep rage peut  tre effectu  en temps r el, par furetage et   l'aide de quelques outils. Nous pr sentons ici les r sultats pr liminaires de nos travaux: une m thode d'alignement de termes complexes dans des textes originaux de diverses langues, un logiciel prototype fonctionnel et une  valuation pr liminaire du prototype sur des textes fran-

(*) D partement de langues et linguistique, Universit  Laval.

çais et anglais appartenant au domaine de l'informatique. Le logiciel intègre les outils suivants: une interface de navigation, un module de reconnaissance automatique des termes complexes (français, anglais et portugais), un module de repérage automatique de termes formellement semblables, un module de génération automatique de réseaux conceptuels (imparfaits mais d'une grande utilité) et un module de repérage des cooccurrents.

MOTS CLEFS: terminologie; équivalence interlinguistique; termes complexes; traduction.

RESUMO: Há já alguns anos, diversos pesquisadores vieram a se interessar pela extração e exploração de conhecimentos contidos em corpúsculos de traduções, entendendo constituir esse um meio de fornecer ao tradutor ferramentas para a identificação de equivalentes interlinguais. A exploração de traduções representa, no entanto, um problema, na medida em que os equivalentes não são redigidos espontaneamente; ora, o primeiro critério que o terminólogo leva em conta é a língua na qual foi redigido o documento que analisa (língua do original ou língua de tradução). Tal fato levou-nos a formular a seguinte indagação: seria possível identificar os equivalentes interlinguais a partir de textos originais em línguas distintas? A resposta é afirmativa. Tomando por base textos redigidos espontaneamente em francês, em inglês e em português, foi possível encontrar de forma rápida, e para um número significativo de termos complexos de uma das línguas, os equivalentes nas duas outras línguas. A única condição a ser respeitada é a utilização de textos que pertençam a um mesmo domínio e que tenham a mesma função comunicativa. A identificação pode ser efetuada em tempo real, com o auxílio de algumas ferramentas. No presente trabalho, apresentamos os resultados preliminares de nossa investigação: um método de alinhamento de termos complexos nos textos originais de diversas línguas, um protótipo de aplicativo funcional e uma avaliação preliminar do protótipo aplicado a textos em francês e em inglês pertencen-

tes ao domínio da informática. O aplicativo incorpora as seguintes ferramentas: uma interface de navegação, um módulo de reconhecimento automático dos termos complexos (francês, inglês e português), um módulo de identificação automática de termos formalmente similares, um módulo de geração automática de redes conceptuais (imperfeitos, mas de grande utilidade) e um módulo de identificação de co-ocorrências.

UNITERMOS: terminologia; equivalência interlingual; termos complexos; tradução.

1. Introduction

Le texte, qui est le véhicule privilégié de l'information, occupe une place importante dans les organisations. L'informatisation importante que nous connaissons depuis plusieurs années a eu pour effet de stimuler considérablement la production documentaire. Par exemple, "chaque année, les entreprises et administrations françaises mettent en circulation (après les avoir rédigées, éventuellement dactylographiées, copiées ou photocopées, archivées), environ quatre cent milliards de pages." (Carré *et al.*, 1991, p. 85) S'il n'y a pas si longtemps on se préoccupait beaucoup de la qualité graphique des documents (impression laser, intégration de tableaux et graphiques, couleur), aujourd'hui on s'intéresse davantage à la gestion et à l'exploitation de ces masses de données textuelles. Tout le monde s'entend pour dire qu'elles représentent une richesse extraordinaire mais encore faut-il être en mesure de l'exploiter efficacement. C'est dans cette perspective que nous travaillons depuis quelques années sur le repérage d'information dans des textes spécialisés.

Plusieurs chercheurs se sont intéressés à l'extraction et à l'exploitation des connaissances contenues dans des corpus de traductions (Brown *et al.*, 1990; Dagan *et al.*, 1991; Gale *et al.*, 1991; Isabelle *et al.*, 1993; Ladouceur, 1991). Ils voyaient là, entre autres, le moyen de fournir au traducteur des outils pour le repérage d'équivalents interlinguistiques. Toutefois, l'exploita-

tion de traductions pose un problème parce que les équivalents ne sont pas rédigés spontanément alors que le premier critère considéré par le terminologue est la langue dans laquelle a été rédigé le document qu'il explore (langue d'origine ou langue de traduction). Cela nous a amenés à nous pencher sur le problème du repérage automatique d'équivalents interlinguistiques à partir de textes plurilingues originaux.

Nos travaux ont d'abord commencé par la création d'un poste de travail pour l'observation et l'analyse, en contexte et hors contexte, des termes complexes contenus dans des textes originaux de diverses langues. Notre objectif était de mettre au point une série d'instruments qui nous permettraient de recueillir des informations sur les équivalents plurilingues qui pourraient ensuite être exploitées pour la conception d'une méthode d'alignement automatique. Dès que les premiers modules de notre poste de travail ont été développés, nous nous sommes aperçus qu'il était possible de repérer assez rapidement, en mode interactif, des équivalents interlinguistiques dans des textes qui appartiennent à un même domaine et qui ont la même fonction de communication si on dispose de quelques outils (un module de reconnaissance automatique des termes complexes, un module de repérage automatique de termes formellement semblables, un module pour la création de réseaux conceptuels et un module pour l'analyse des cooccurrents) et d'une bonne méthode de recherche.

2. Méthode d'alignement de termes complexes extraits de textes originaux

L'approche préconisée pour l'alignement des termes complexes est fondée sur deux observations :

1. Tous les termes complexes d'un domaine peuvent être reliés les uns aux autres soit directement, lorsqu'ils ont un ou plusieurs mots graphiques communs ou qu'ils ont un ou plusieurs cooccurrents communs (ils sont alors plus ou moins sémantiquement apparentés), soit indirectement lorsqu'ils

appartiennent à des réseaux conceptuels qui ont un terme commun ou à des réseaux conceptuels qui sont reliés (par un terme commun) par l'intermédiaire d'un ou plusieurs réseaux conceptuels.

carte mère	application graphique	interface utilisateur
carte à microprocesseur	carte graphique	interface monochrome
carte de son	graphique 3D	interface scsi
carte à puce	interface graphique	interface edi
carte graphique		interface graphique

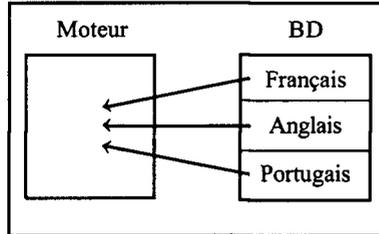
2. Des termes appartenant au même domaine mais à des langues différentes peuvent être reliés s'ils sont formellement presque identiques (il est fort probable qu'ils désignent la même chose) ou s'ils ont en commun un ou plusieurs mots graphiques qui ont une forte ressemblance formelle (ils font parties de réseaux conceptuels dont l'un contient fort probablement au moins un équivalent interlinguistique d'un terme de l'autre réseau).

application graphique	3D graphic software
carte graphique	3D graphic accelerator
graphique 3D	graphics controller
interface graphique	graphics viewer

Sur la base de ces observations nous avons élaboré une méthode d'alignement de termes complexes qui comporte deux phases. La première est un pré-traitement automatique qui consiste à reconnaître les termes complexes dans les textes et à en aligner un certain nombre sur la base de leur ressemblance formelle. La deuxième phase est une séance de furetage (donc interactive) où l'utilisateur tente de repérer un équivalent en parcourant les liens qui existent entre les termes complexes d'une langue et ceux des autres langues. Cette phase de furetage est assistée de fonctions spécialement conçues pour l'alignement.

2.1 Reconnaissance automatique des termes complexes

Pour la reconnaissance automatique des termes complexes, nous utilisons le logiciel d'extraction terminologique **Term-plus** (Ladouceur et Cochrane, 1996) qui a été développé au CIRAL, à l'Université Laval. **Term-plus** a été conçu dans une perspective de plurilinguisme. Il comporte un moteur d'analyse et une banque de données linguistiques pour chaque langue qu'il traite. Jusqu'à maintenant, nous avons préparé les banques de données pour le français et l'anglais. Nous travaillons présentement sur la banque du portugais¹. Voici des exemples de résultats que nous obtenons à partir de trois revues (française, anglaise et portugaise) d'informatique:



Français	Anglais	Portugais
carte graphique	active-matrix	banco de dados
cd-rom	cd-rom	cd-rom
disque dur	color 500	cds interativos
fonctions vocales	digital camera	leitora de cds
lecteur de cd-rom	fax modem	leitura sequencial
matrice active	floppy disk	placa de som
mo de mémoire	hard disk	processador de textos
pc multimédia	internet explorer	tempo de acesso
service lecteurs	pentium pro	tempo real
story board	third-party	unidades de cd
windows news	visual basic	unidades de cd-rom
windows nt	website builder	windows 95
		windows nt

(1) Nous avons effectué des tests sur le portugais avec une banque de données presque vide et avons obtenu des résultats tout de même suffisamment intéressants pour intégrer le portugais à nos expériences d'alignements.

Ces résultats ne sont pas parfaits. Ils contiennent des éléments qui ne sont pas des termes complexes. Il faut dire que les deux tiers des résultats que **Termplus** produit sont du bruit² (les listes de notre exemple ont bien sûr été épurées). Cette quantité de bruit est le prix que nous acceptons de payer pour éviter les silences³. Si **Termplus** retient beaucoup de segments erronés par contre il n'oublie à peu près pas de termes. On remarque également que dans notre exemple le terme *cd-rom* apparaît dans les trois listes. Un simple module de comparaison peut rapidement établir des liens entre ces trois mots.

2.2 Établissement des premiers liens

Parce que l'emprunt interlinguistique et intralinguistique est un phénomène relativement fréquent en langue de spécialité (et particulièrement dans le domaine de l'informatique), nous retrouvons, dans des textes de langues différentes, plusieurs termes identiques ou qui ont une forte ressemblance formelle : *cd rom / cd rom, matrice active / active matrix*. Plus le niveau de ressemblance formelle est élevé entre deux termes (qui proviennent de textes de langues différentes), plus la probabilité est forte que ces termes désignent un même objet ou un même concept.

Après avoir isolé les termes complexes extraits des différents textes, nous pouvons établir très facilement une première série de liens à l'aide d'un calcul de similitude. Il existe plusieurs façons de calculer le niveau de ressemblance entre des mots. La méthode que nous avons adoptée consiste à former une série de doublons (deux caractères consécutifs) pour les mots que nous voulons comparer. Par exemple, avec *matrice active* et *active matrix* nous obtenons:

ma, at, tr, ri, ic, ce, ac, ct, ti, iv, ve / ac, ct, ti, iv, ve, ma, at, tr, ri, ix.

Nous calculons ensuite le pourcentage des doublons de la première série qui apparaissent également dans la seconde sé-

(2) Éléments retenus mais qui ne sont pas des termes.

(3) Termes qui auraient dû être retenus mais qui ont été omis.

rie, nous faisons l'inverse et nous calculons finalement la moyenne des deux. Avec notre exemple, nous obtenons un score de 86%. La méthode est simple, facile à implémenter et donne d'excellents résultats.

2.3 Fonctions de furetage

La suite du traitement s'effectue en temps réel, au moment où l'équivalent dans l'autre langue est recherché. Nous proposons une méthode de furetage assistée de trois fonctions: extraction des termes formellement semblables, construction de réseaux conceptuels⁴ et analyse des cooccurrents.

2.3.1 Balayage des termes formellement semblables

En mode automatique, les seuls termes qu'il est prudent d'aligner sur la base de leur ressemblance (sinon le taux d'erreur serait évidemment trop élevé) sont ceux qui sont presque identiques c'est-à-dire qui ont un pourcentage de similitude égal ou supérieur à 85%. Ceci étant dit, il ne faut pas conclure qu'il n'y a pas de termes avec un pourcentage de similitude inférieur à 85% qui devraient être alignés. En affichant les termes qui ont une cote de similitude se situant entre 60% et 85%, nous obtenons une liste bruyante mais que l'utilisateur peut réviser assez rapidement. Il peut ainsi repérer et aligner en quelques secondes (voire une minute ou deux si la liste est particulièrement longue) des termes comme :

documentation interactive / online documentation (60%)
navigation aisée / easy navigation (60%)
interface graphique / graphical interface (70%)
modem externe / external modem (70%).

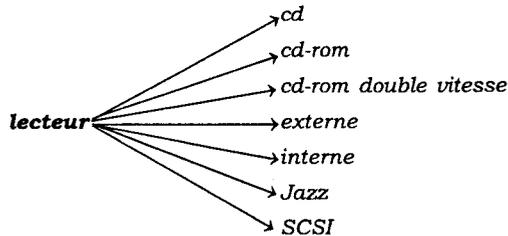
Si on abaissait le seuil en deçà de 60%, on obtiendrait alors des listes très bruyantes et l'utilisateur perdrait beaucoup de temps à les réviser.

(4) Les réseaux conceptuels produits automatiquement par notre prototype sont très imparfaits mais forts utiles.

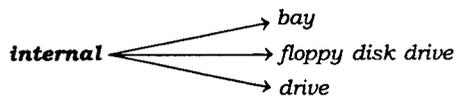
2.3.2 Affichage de réseaux conceptuels

La création d'un réseau conceptuel consiste à regrouper dans une liste tous les termes complexes qui contiennent un mot donné. Nous parlerons alors du réseau conceptuel de *lecteur* pour désigner la liste des termes complexes qui contiennent le mot *lecteur*. Nous obtenons donc un sous-ensemble bruyant de ce que serait un véritable réseau conceptuel. Parce que nous travaillons en temps réel, nous devons privilégier une technique rapide même si elle est imparfaite.

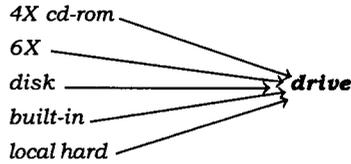
L'affichage de réseaux conceptuels, combiné au repérage de mots formellement semblables, est d'après nous l'outil de furetage par excellence. Par exemple, si l'utilisateur cherche un équivalent anglais pour *lecteur intégré*, il ne pourra pas l'obtenir uniquement par le biais d'une analyse de la ressemblance formelle des termes. Mais, s'il fait afficher le réseau conceptuel de *lecteur*, il obtient ceci:



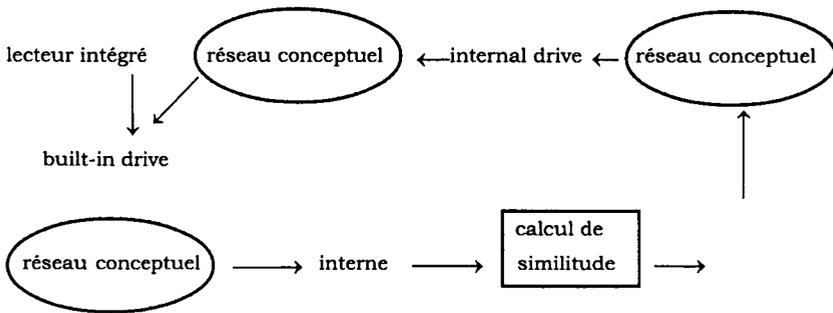
Parmi ces termes, il repérera *lecteur interne* qui a un sens assez près du terme source. Ensuite, le mot *interne* peut être utilisé comme pont (manuellement ou automatiquement par une analyse de similitude) vers le terme anglais *internal*. Si l'utilisateur fait ensuite générer le réseau conceptuel de *internal*, il obtient cette liste de termes:



Il repère ainsi l'équivalent anglais de *lecteur interne* soit *internal drive*. Finalement, il peut faire afficher le réseau conceptuel de *drive* et obtient:



dans laquelle apparaît l'équivalent recherché soit *built-in drive*. Pour retrouver son équivalent anglais, l'utilisateur a donc parcouru le chemin suivant:

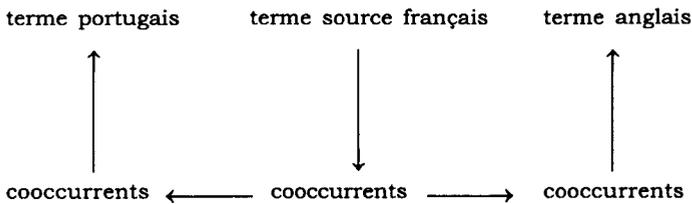


Évidemment, si l'utilisateur connaît dès le départ l'équivalent anglais de *lecteur* c'est-à-dire *drive*, il peut faire générer directement le réseau conceptuel construit autour de *drive* et trouver beaucoup plus rapidement l'équivalent anglais recherché.

2.3.3 Analyse des cooccurrents

Si l'équivalent d'un terme n'a pu être trouvé, l'utilisateur peut tenter d'établir des ponts avec les autres textes par le biais des cooccurrents. L'idée est que lorsqu'il n'est absolument pas possible d'établir de lien, par la ressemblance formelle directe

ou indirecte (par le biais des réseaux conceptuels), entre des termes qui sont des équivalents interlinguistiques comme *caixas de som* / *haut-parleurs* / *speakers*, on peut toujours tenter de passer par les cooccurrents. Nous utilisons le mot *passer* car le but est littéralement de découvrir un passage entre deux textes, passage qui nous amènera le plus près possible du terme recherché. L'utilisateur peut, par exemple, faire afficher les cooccurrents les plus fréquents et tenter de trouver des mots formellement semblables dans les autres textes et à partir de ces mots repérer leurs cooccurrents pour tenter de trouver l'équivalent du terme source :



Par exemple, l'un des cooccurrents du terme portugais *caixas de som* est *microfone* qui peut être relié (par ressemblance formelle) au mot (français, et anglais) *microphone*. Ce dernier conduit alors au cooccurrent *haut-parleurs*, en français, et *speakers*, en anglais. Même si nous commençons tout juste à explorer l'analyse des cooccurrents nous croyons que cette voie est très prometteuse.

3. Pluritermes: un logiciel d'alignement interactif

Pluritermes est un logiciel prototype. Il n'offre pas encore toutes les fonctionnalités d'un logiciel commercial. Par exemple, il ne permet pas à l'utilisateur d'afficher le contexte où figure le terme complexe, ce qui est indispensable dans bien des cas pour déterminer le sens précis d'un terme. Dans le cadre de notre expérimentation, nous devons recourir à la fonction recherche d'un logiciel de traitement de texte pour faire afficher le contexte de certains termes. De plus, il ne conserve aucune

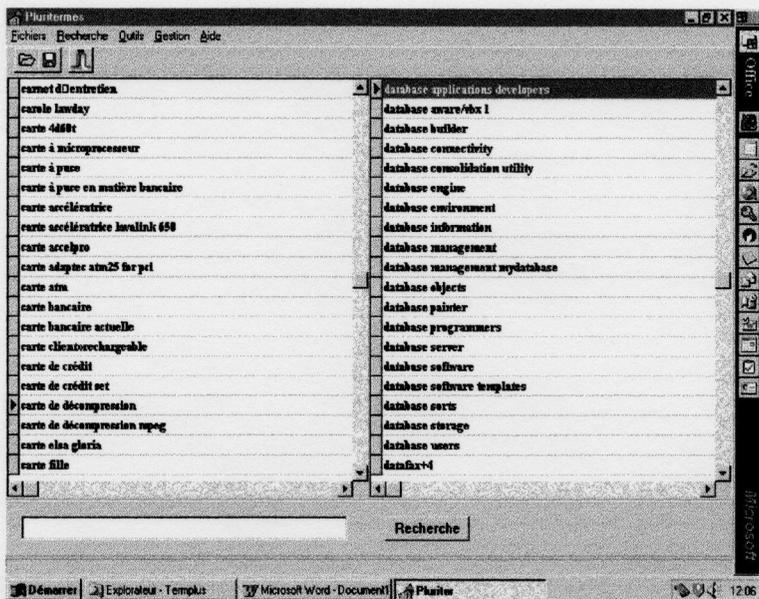
trace du cheminement effectué pour repérer un équivalent interlinguistique. Le principe même du furetage implique un processus de parcours permettant des rétroactions. Dans cette optique, il serait particulièrement utile de conserver en mémoire non seulement le trajet parcouru, mais également tous les réseaux conceptuels et les listes de cooccurrents générés pendant le processus complet de repérage.

Pluritermes présente à l'utilisateur deux fenêtres qui contiennent chacune une liste de termes complexes générée par **Termplus**. L'utilisateur peut parcourir directement l'une ou l'autre des listes pour y rechercher un terme. Il peut également utiliser un module de repérage pour effectuer une recherche dans une des deux listes ou dans les deux listes simultanément (voir écran 1).

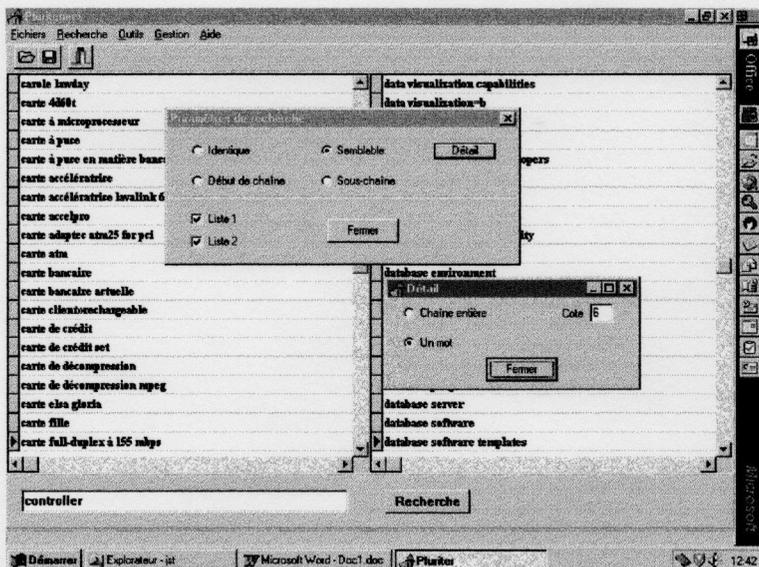
L'engin de recherche de **Pluritermes** peut être paramétré (voir écran 2). Il permet ainsi à l'utilisateur de rechercher:

- un terme complexe précis,
- tous les termes complexes qui commencent par un mot en particulier,
- tous les termes complexes qui contiennent un mot en particulier (c'est ce mode de recherche qui permet de générer un réseau conceptuel),
- un terme complexe formellement semblable à un terme donné (le niveau de ressemblance peut être lui-même paramétré),
- tous les termes complexes qui commencent par un mot formellement semblable à un mot donné,
- tous les termes complexes qui contiennent un mot formellement semblable à un mot donné.

Écran 1



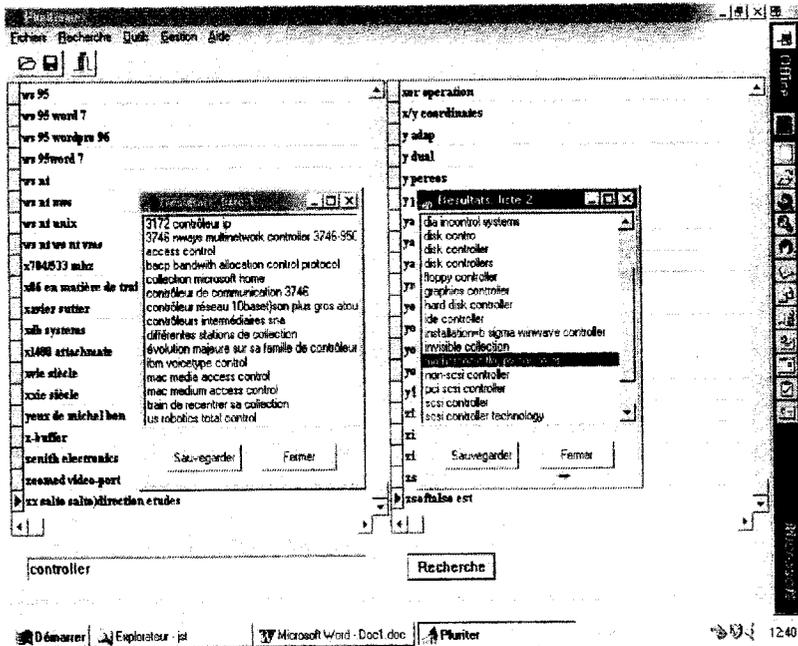
Écran 2





Lorsque l'utilisateur veut générer un réseau conceptuel, il effectue une recherche à partir d'un mot et il paramètre l'engin de recherche pour qu'il repère tous les termes complexes qui contiennent ce mot ou un mot formellement semblable à ce mot. Le résultat apparaît dans une nouvelle fenêtre, ou dans deux fenêtres si la recherche a été effectuée sur les deux listes simultanément (voir écran 3).

Écran 3



4. Évaluation

Les performances d'un logiciel comme **Pluritermes** dépendent de plusieurs facteurs. Tout d'abord, il faut que les textes utilisés appartiennent au même domaine de spécialité, qu'ils traitent de sujets semblables et qu'ils aient la même fonction de

communication. Deux revues d'informatique par exemple qui portent sur Windows 95 et qui ont été publiées à peu près en même temps peuvent donner d'excellents résultats. La taille des textes est très importante et tout particulièrement celle du texte cible. Évidemment, plus la taille du texte cible est grande plus il y a de chances que nous y retrouvions l'équivalent interlinguistique recherché.

L'expérience de l'utilisateur joue également un rôle non négligeable. L'habilité de l'utilisateur à repérer un équivalent interlinguistique augmente avec la pratique. On observe d'ailleurs le même phénomène chez les internautes qui recherchent des informations sur Internet. La méthode de recherche que l'utilisateur emploie est également importante. Celle que nous proposons a été élaborée de façon intuitive. Nous procédons actuellement à une série d'expériences de repérage avec plusieurs utilisateurs sur une série de paires de textes appartenants à des domaines différents afin d'élaborer une méthode de recherche plus rigoureuse. Nous travaillons avec des textes de taille variée et des paires de textes qui portent sur des sujets plus ou moins semblables. Nous avons également choisi nos utilisateurs en fonction de la connaissance qu'ils ont de l'anglais et du français (langues sur lesquelles porte notre évaluation) afin de mesurer l'impact de la connaissance des langues source ou cible sur la qualité des résultats et de la vitesse de repérage.

En raison de sa complexité, l'évaluation précise et objective de **Pluritermes** reste à venir. Mais, voici tout de même le fruit de nos premières observations. Tout d'abord, il faut préciser que notre prototype n'est pas très rapide. Par exemple, la génération de deux réseaux conceptuels (un par langue) à partir d'un mot semblable à un mot donné prend environ deux minutes et demi. Cela signifie que nous passons beaucoup de temps à attendre que le logiciel ait terminé une phase de traitement. Cette attente fausse nos observations car la plupart du temps, entre deux phases de traitement, s'écoule un temps très bref (une seconde ou deux). On pourrait donc croire qu'en rendant par exemple le logiciel deux fois plus rapide que la recherche nécessiterait deux fois moins de temps. Cela est tout à fait faux

puisque pendant les deux minutes et demi d'attente, l'utilisateur peut réfléchir à la suite des événements. Si on fait abstraction du temps machine, le temps nécessaire pour repérer un équivalent interlinguistique est à peu près le temps de réflexion multiplié par le nombre de phases de traitement. Mais le temps de traitement peut varier beaucoup selon l'état d'avancement de la recherche. Bref, il ne faut pas extrapoler trop rapidement à partir de nos observations.

La plupart des termes que nous avons recherchés (nous avons toujours travaillé du français vers l'anglais), et qui étaient dans le texte cible, ont été trouvés assez rapidement et n'ont nécessité que quatre ou cinq phases de traitement. Si on ne tient pas compte du temps de traitement cela signifie entre une et deux minutes par terme. Autant que possible, nous recherchions des termes pour lesquels nous savions que les équivalents étaient formellement très différents parce que tous les termes formellement semblables comme **interface graphique** et **graphical interface** sont repérés du premier coup. Par exemple, avec un terme comme **carte mère** nous avons trouvé **motherboard** en quatre étapes :

1. le repérage de termes formellement semblable n'a rien donné (évidemment);
2. à partir du réseau conceptuel de **carte**, nous avons retenu **graphique** dans **carte graphique**;
3. à partir des réseaux conceptuels anglais et français construits à partir des mots formellement semblables à **graphique**, nous avons retenu **graphical controller** (du réseau anglais);
4. à partir du réseau conceptuel construit à partir de **controller** (cette fois-ci uniquement à partir de la liste de termes anglais), nous avons repéré le terme **on-the-motherboard floppy controller** qui contenait le terme recherché soit **motherboard**.

5. Applications de Pluritermes pour les traducteurs, les terminologues et les rédacteurs

L'importante augmentation de la demande de services terminologiques rend **Pluritermes** d'autant plus utile. Selon Gouadec, cela est attribuable à quatre facteurs : l'accélération des transferts techniques, économiques, commerciaux, technologiques et industriels, la poussée de la demande de services linguistiques, la définition d'une politique linguistique cohérente et l'effet terminotique-automatique (1993, p. 5-8). L'Union européenne⁵ est sans doute le plus bel exemple en ce qui concerne les besoins de services linguistiques. En effet, l'Union, qui se compose de 12 États membres et dont la population atteint 335 millions d'habitants, a neuf langues officielles: l'allemand, l'anglais, le danois, l'espagnol, le français, le grec, l'italien, le néerlandais et le portugais. Toutes ces langues sont traitées sur un même pied d'égalité, ce qui signifie que les documents produits dans l'une ou l'autre des langues officielles doivent être obligatoirement traduits dans les huit autres langues. Par conséquent, l'informatisation partielle de la tâche des traducteurs est l'une des priorités de la Commission européenne.

Il semblerait d'ailleurs que le recours à des outils informatiques fasse épargner entre 15 et 20 % du temps au traducteur (voir Roburin, 1989 cité par Gaudin, 1993, p. 124). Cette économie est bienvenue d'autant plus que chaque minute du traducteur est comptée s'il veut répondre aux demandes sans cesse croissantes et respecter les délais toujours très serrés tout en offrant une traduction de haute qualité dans les différentes sphères d'activités.

Au cours des dernières années, les besoins des traducteurs ont changé, c'est pourquoi nous devons adapter les produits qui leur sont destinés afin de les aider à accroître leur productivité au lieu de leur faire perdre temps et énergie. Les dictionnaires et fiches bilingues où seuls les termes hors con-

(5) Les renseignements relatifs à la situation qui prévaut dans l'Union européenne sont tirés de l'article de Barbara Stork (1995, pages 195-210).

texte apparaissent sont aujourd'hui dépassés. Le traducteur a besoin d'outils fiables et performants qui lui ouvriront toutes grandes les portes de l'information. Dancette souligne à cet effet que «les dictionnaires seraient beaucoup plus utiles et plus fiables pour la traduction s'ils contenaient une composante conceptuelle riche et structurée» (1995, p. 162).

5.1 Utilité de Pluritermes

Avec **Pluritermes**, notre objectif est de faciliter l'organisation et la classification des connaissances, en allant au-delà de la simple désignation des notions dans les différentes paires de langues. Plus précisément, nous cherchons à définir avec le plus de précision et de rigueur possible les relations sémantiques, hiérarchiques et associatives qui existent entre les termes. Il importe donc de prendre en compte, outre le terme, les éléments qui l'entourent (les tournures nominales, verbales et adjectivales ainsi que les termes connexes) et qui constituent le réseau notionnel. Ce réseau permet entre autres de relever, en contexte réel d'utilisation, les diverses formes employées en langue technique et scientifique, leurs significations ainsi que leurs équivalents.

L'utilité la plus manifeste de **Pluritermes** est l'élaboration de réseaux sémantiques et notionnels propres à un domaine ou à un sous-domaine. Comme nous l'avons déjà mentionné, le logiciel permet à l'utilisateur d'ouvrir, à partir d'une liste de paires de termes potentiellement équivalents, des réseaux conceptuels et de fureter çà et là à son gré.

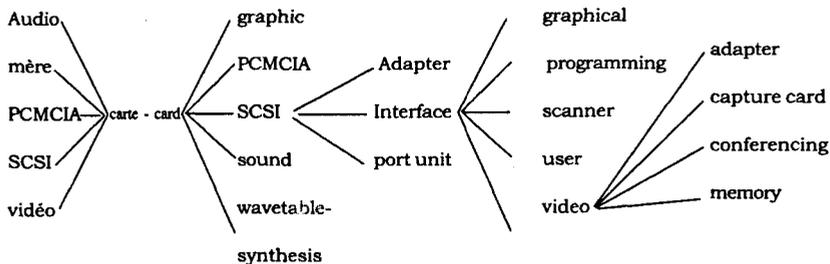
5.2 Exploitation des réseaux sémantiques et notionnels

Nous avons représenté, dans la figure ci-dessous, les différentes étapes de l'analyse effectuée par le logiciel. Après avoir inséré dans le logiciel des textes parallèles⁶ traitant d'informatique, **Pluritermes** a produit une liste exhaustive de paires de

(6) Nous entendons par textes parallèles, des textes qui traitent d'un même sujet, mais que les auteurs ont rédigés indépendamment dans leur langue respective.

termes. Nous avons retenu, pour les besoins de la démonstration, les équivalents proposés *carte - card* et avons décidé de fureter dans le texte anglais. Examinons d'un peu plus près comment s'effectue ce furetage et ce que nous pouvons trouver en cours de route.

Le principe de base du furetage ressemble à celui employé dans Internet, c'est à dire que l'utilisateur n'a qu'à cliquer sur le mot qui l'intéresse pour que toute l'information relative à ce mot apparaisse à l'écran. Dans notre cas, le logiciel ouvre des fenêtres (une par langue) et présente la liste des termes retenus dans les textes. C'est grâce aux menus déroulants synchronisés avec le texte à l'étude que l'utilisateur peut voir les mots en contexte. Dans notre exemple, le terme *card* apparaît à 66 reprises dans le texte. Si l'utilisateur ne veut pas consulter toutes les occurrences du mot, il peut restreindre le domaine à un sous-domaine précis pour n'obtenir qu'un seul type de carte, par exemple *SCSI card*. Le logiciel indique alors qu'il y a dans ce texte une seule occurrence du terme recherché et la présente en contexte. L'utilisateur peut alors être tenté de regarder quels sont les autres termes qui sont composés du mot SCSI et ainsi de suite jusqu'à ce qu'il trouve un mot qui appartenait à la première série de termes recherchés, c'est-à-dire *video capture card*. Ce fil d'Ariane (représenté par les traits en gras) permet à l'utilisateur d'avoir accès à un réseau complexe de connaissances relatives au concept, aux termes qui l'expriment dans chacune des langues, ainsi qu'aux relations qui lient ce concept aux concepts voisins par quelque critère que ce soit. Si on reprend l'exemple de *SCSI card*, l'utilisateur intéressé par des équivalents en portugais trouvera rapidement *placa SCSI*.



Il importe de retenir que les données sont présentées comme des suggestions afin d'aider l'utilisateur (terminologue, traducteur ou rédacteur) à comprendre le sens, à instaurer des usages, éventuellement même à enrichir et mettre à jour des dictionnaires là où ils sont incomplets, quant aux équivalents interlinguistiques ou à l'information conceptuelle. Si le terme n'est pas défini, le recours aux contextes où ils apparaissent permet à l'utilisateur d'en inférer le sens. Il lui est ainsi possible de découvrir des réseaux insoupçonnés d'information dans l'une ou l'autre des langues, car les liens notionnels sont tout aussi importants que les notions elles-mêmes.

Cette méthode peut également aiguiller le traducteur ou le rédacteur vers les tournures à privilégier dans son texte. Par exemple, une recherche exploratoire dans les trois revues d'informatique a fait ressortir que pour le nom *application*, les verbes français employés sont *concevoir, convertir, créer, développer, exécuter, exploiter, faire dialoguer, faire tourner, fermer, installer, lancer* et *réaliser*, alors qu'en anglais ce sont les verbes *to build, to develop, to execute, to exit, to launch, to move, to set up* et *to register* qui s'apparient avec ce nom. En portugais ce sont les verbes *executar, rodar* et *acessar* qui sont retenus.

Ces renseignements fort utiles au moment de rédiger un texte sont quasi-inexistants dans les dictionnaires bilingues ou multilingues. Quant aux grandes banques de terminologie (BTQ et Termium), elles ne consignent que les usages des termes en contexte ainsi que leur définition. Toutefois, l'ouverture des marchés et la polyvalence exigée des traducteurs et des rédacteurs obligent à revoir la consignation des renseignements dans les ouvrages de référence. Il faut aller bien au-delà des dictionnaires de locutions idiomatiques pour s'intéresser plus précisément aux phraséologismes propres aux langues de spécialité.

Comme le souligne d'ailleurs à juste titre Dubuc: "Dans sa sélection d'éléments lexicaux significatifs, le terminologue ne doit pas s'en tenir aux seuls concepts de base: sa recherche doit englober tout l'arsenal des moyens d'expression, oraux ou écrits, du domaine étudié, y compris les verbes et locutions verbales ou adjectives qui appartiennent en propre à ce domaine" (1992, p. 3). Seule cette façon de faire peut permettre à la termi-

nologie de se tenir en résonance avec l'usage afin de mieux répondre aux besoins des utilisateurs.

En confrontant les contextes prélevés, chacun pourra dégager sans *a priori* tous les emplois possibles des mots dans chacune des langues. Ainsi le terminologue trouvera des informations qui lui permettront d'enrichir les banques de données, d'établir des arbres de domaines, et de fournir des attestations d'emploi. Pour leur part, le traducteur et le rédacteur auront à leur disposition des exemples d'emploi en contexte et des tournures idiomatiques.

Bibliographie

- BROWN, P. *et al.* (1990) A statistical approach to language translation. *Computational Linguistics*, v. 16, n. 2.
- CARRÉ, R. *et al.* (1991) *Langage humain et machine*. Presses du CNRS, Paris.
- DAGAN, I., ITAI, A. et SCHWALL, A. U. (1991) Two languages are more Informative than one. *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, Berkeley.
- DANCETTE, J. (1995) Organisation conceptuelle du domaine et structure de dictionnaire – L'exemple du commerce au détail. In: BOU-LANGER, Jean-Claude (dir.). *TTR (Technolectes et dictionnaires)*, v. VIII, n. 2.
- DUBUC, R. (1992) *Manuel pratique de terminologie*. Montréal, Linguatex, XIII.
- GALE, W. A. et CHURCH, K. W. (1991) A program for aligning sentences in bilingual corpora. *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*. Berkeley.
- GAUDIN, F. (1993) *Pour une socioterminologie, Des problèmes sémantiques aux pratiques institutionnelles*. Rouen, Publications de l'Université de Rouen, n. 182.
- GOUADEC, D. (1990). *Terminologie, constitution des données*. Paris, AFNOR, XX.
- ISABELLE, P. et WARWICK-ARMSTRONG, S. (1993) Les corpus bilingues: une nouvelle ressource pour le traducteur. In: BOUILLON, P. et CLAS, A. (dir.) *La traductique*. Montréal, Presses de l'Université de Montréal.

- LADOUCEUR, J. (1991) La fusion de textes bilingues. *Terminologie et documentation dans la communication spécialisée*. Hull.
- LADOUCEUR, J. et COCHRANE, G. Termpus: un logiciel d'extraction terminologique. *Terminologies nouvelles n. 15*, Agence de coopération culturelle et technique et la Communauté française de Belgique, juin et décembre 1996.
- PC MAGAZINE BRASIL (1996), São Paulo, Editorial América do Brasil, 120 p.
- STORK, B. (1995), Les langues dans l'Union européenne: le travail des traducteurs et des interprètes. In: AUBIN, M-Ch. (dir.) *Perspectives d'avenir en traduction*. Winnipeg, Presses universitaires de Saint-Boniface.
- WINDOWS NEWS (1996), n. 35, France, Edicorp.
- WINDOWS MAGAZINE (1996), v. 7, n. 10, New York, CMP Media Inc.

