

Conclusion

Certains de nos objectifs ont été atteints, d'autres sont perfectibles en continuant le travail, d'autres encore sont à l'état de perspectives.

Au départ de ces travaux, nos objectifs étaient d'apporter des solutions novatrices pour participer aux recherches qui tendent à transformer la machine en un média à valeur ajoutée. Cette valeur ajoutée concerne l'accès personnalisé à des documents et l'exploration assistée de leur contenu. La structure en dispositif apporte la possibilité de décrire des centres d'intérêt particuliers de façon corrélée à une tâche et par là-même de tenir compte de dimensions particulières de la situation d'utilisation du système. L'approche différentielle et componentielle permet de contraindre la structuration des données pour éviter une explosion quantitative et les rendre utilisables par des processus automatiques. Ces contraintes ont également pour effet d'assister l'utilisateur dans l'expression de ces besoins. Les limites rencontrées par l'utilisation de ressources et de traitements légers peuvent se combler ou se résoudre en partie par l'interaction, par une négociation entre les usagers, la machine et le terrain commun que représentent les dispositifs et les informations qui leurs sont associées. Les principes de *sémantique légère* pour le TAL montrent leur intérêt lorsque l'utilisateur est au cœur du système.

Si les fondements de nos travaux se trouvent dans la Sémantique Interprétative, nous avons pu voir tout au long de ce tapuscrit que nous nous sommes petit à petit écarté des concepts de cette théorie pour finalement en créer de nouveaux : ceux de notre modèle computationnel. Au final, s'il peut s'avérer difficile pour les habitués de cette approche de « retrouver leurs petits », c'est avant tout parce que la confrontation avec une réalité computationnelle et applicative amène nécessairement des changements plus ou moins importants. C'est, à notre avis, l'un des intérêts des échanges interdisciplinaires.

Les logiciels implantés à partir des principes de LUCIA sont des logiciels d'étude (c.f. note 137). Ils doivent être adaptés pour des utilisations réelles. Les perspectives de ce travail sont donc nombreuses. On peut distinguer des perspectives théoriques et des perspectives applicatives, tout en gardant à l'esprit que les unes conditionnent évidemment les autres. Parmi les perspectives théoriques, il peut être intéressant de justement revenir aux sources de ces travaux et d'examiner plus précisément quelles sont les modifications conceptuelles apportées aux principes de la SI. Par exemple, l'étude sur la métaphore a montré qu'il était possible avec LUCIA de retrouver la nature créatrice et analogique du lien métaphorique. Il peut être envisagé de maintenant examiner les différences de nature entre cette analogie et celle abordée par les

autres travaux ayant pour objet le même fait de langue. Parmi les perspectives applicatives, il est évident que des travaux communs avec des spécialistes de l'acquisition terminologique ou de connaissances à partir de documents permettraient d'améliorer nos propositions actuelles. Mais il est envisageable également d'augmenter encore la place de l'individu dans le système et de travailler maintenant sur la possible constitution des dispositifs à partir d'autres types d'interactions. Nous pensons en particulier au dialogue en langue naturelle entre l'utilisateur et la machine. Des travaux ont déjà montré leur intérêt dans ce champ de recherche. Ceux exposés dans [Pohl *et al.*, 1995] relatifs à l'analyse de présuppositions pour établir des hypothèses sur un interlocuteur humain sur la base du type d'acte de la parole qu'il utilise, nous paraissent possiblement utilisables. De même, certains aspects du travail de Beust [Beust, 1998*] vont dans ce sens et pourraient constituer une base de recherche intéressante avec des partenaires universitaires spécialistes du dialogue Homme/Machine que nous côtoyons déjà (comme l'équipe du laboratoire LIUM de l'université du Mans). Beust a effectivement montré que dans de nombreux échanges langagiers, dès qu'une imprécision ou une ambiguïté empêchait leur bon déroulement, c'était l'expression des différences entre les termes qui faisaient sens. L'expression de ces différences pourrait alors être la base d'une construction de dispositif.

Nos travaux sont également à l'origine d'une nouvelle thèse qui vient de débiter. Thibault Roy travaille actuellement sous la co-direction de Jacques Vergne et Pierre Beust du GREYC sur la visualisation cartographique d'ensembles documentaires et la caractérisation du contenu de documents à l'aide de LUCIA. Ces travaux ont déjà fait l'objet d'une publication [Roy et Beust, 2004*].

Nous avons proposé dans cette thèse un modèle pour l'accès personnalisé aux documents et l'exploration assistée de leurs contenus. Ce modèle prend place dans les recherches en informatique qui visent à améliorer la machine en tant que média à valeur ajoutée pour les tâches documentaires. Plus largement, il s'inscrit dans les champs de recherche des STIC (Sciences et Technologies de l'Information et des communications).

Les STIC continuent de bouleverser le monde avec une rapidité déconcertante. À l'heure où sont écrites ces lignes, l'ordinateur est devenu un outil de propagande au service de la terreur et des combats les plus louables, au service du commerce et des échanges amoureux, au service de tout à chacun pour peu que l'ordinateur à disposition soit connecté à un autre. La mémoire collective, le travail, les connaissances, toutes ces notions qui remontent à l'aube de l'humanité ont connu de profonds bouleversements avec l'arrivée de l'ordinateur. Les informaticiens, les linguistes, les sociologues, les sociolinguistes, les historiens, les cognitivistes, les pédagogues... tous ont participé à l'élaboration des concepts et des objets qui sont aujourd'hui les supports de communication et de travail de millions d'individus. Ces transformations

technologiques ne sont pas achevées. Il serait hasardeux de se risquer à quelques prospectives que ce soit sur le devenir des travaux des STIC. Les physiciens arrivent chaque jour un peu plus à consolider la place de Moore au panthéon des génies du siècle dernier : les performances des machines ne cessent de s'accroître tandis que leur taille et leur coût décroissent. Mais les attentes des utilisateurs augmentent tout aussi rapidement. Pour traiter des textes, l'engouement suscité par les progrès techniques a depuis quarante ans été au cœur d'un paradoxe. C'est parce que des travaux ont permis de constater notre impossibilité a priori à délimiter les connaissances utiles pour traiter du sens des textes dans une situation donnée, que l'on a cherché à mettre dans la machine toutes les connaissances que l'on pouvait y placer. Parallèlement, des recherches qui s'intéressaient à l'individu, ou aux groupes sociaux dans leurs particularités, montraient que finalement, chaque texte, chaque document, avait un intérêt singulier et pouvait susciter un apport particulier en fonction de la situation de sa lecture et du lecteur lui-même. Dans notre modèle, la compétence des logiciels se constitue au fur et à mesure des usages par un processus itératif. Ces itérations permettent également à l'utilisateur de structurer son savoir.

Mais LUCIA ne permet pas de représenter la connaissance en soi, la connaissance a priori ou la connaissance absolue, universelle. Ce qui est représenté dans les structures du modèle, ce sont les différences entre les connaissances, ce qui fait différence pour une tâche située, pour un utilisateur ou un groupe d'utilisateur donné. En ce sens, notre travail propose une transposition d'une partie de l'héritage de la linguistique saussurienne dans le domaine de la représentation des connaissances. Si pour exprimer les différences nous nous sommes basé sur des attributs originellement affiliés aux sèmes de la Sémantique Interprétative, il est parfaitement envisageable de pouvoir les représenter en machine par d'autres symboles autant signifiants pour les hommes et autant distinguables par les logiciels. Nous l'évoquions dans le chapitre 3 (p.79), les valeurs de nos attributs pourraient prendre la forme de ☺ ou de ☹, ou de ✨, ⚡ et de ☄ pour autant que des agrégats de ces signes pourraient exprimer des différences pour les usagers qui les choisiraient, pour autant que les entités lexicales décrites et catégorisées par ceux-ci soient repérées dans des textes et manipulées automatiquement pour rendre des services également appréciables. Le rapport entre la machine et le sens des textes n'est pas le même que celui entre lecteur humain et les textes. Cependant, nous savons que les logiciels peuvent faire des différences, repérer des récurrences, faire des rapprochements pour peu qu'on leur donne quelques règles de conduite.

Les connaissances manipulées dans notre modèle sont supportées en partie par des entités lexicales. La constitution de ressources lexicales n'est pas réservée aux approches automatisées : ce tapuscrit en apporte les preuves. Partant de ce constat, Pincemin [Pincemin, 1999b*] s'interroge sur le devenir du

traitement automatique dans ce champ d'étude : l'économie de temps qui le caractérise¹³⁸ amènera-t-il tôt ou tard à remplacer le travail de l'expert ? Ou inversement, la qualité des réalisations mobilisant l'intelligence et l'expertise humaine, auxquelles un logiciel ne saurait se substituer, condamne-t-elle à la médiocrité les efforts des traitements automatiques ? Pour Pincemin, le débat ne se situe pas là, car ces démarches sont complémentaires. Des travaux comme ceux décrits dans [Habert, 1998] et [Habert et Nazarenko, 1996*] proposent une troisième voie qui conjugue les approches manuelles et automatisées en utilisant le calcul comme aide à l'expert pour sa tâche de construction de classes sémantiques. Les travaux présentés dans ce tapuscrit en proposent une quatrième où les processus automatiques ne sont pas nécessairement là pour assister un expert, mais peuvent assister un simple usager. Il n'est bien sûr pas envisageable de transformer tous les utilisateurs des ordinateurs en des experts, et encore moins en des experts terminologues ou lexicologues, sous prétexte qu'ils désirent un accès personnalisé au matériau textuel. Si les résultats obtenus de nos propositions sont satisfaisants sur de nombreux plans, ils ne souffrent pas encore la comparaison avec des systèmes où le travail est en grande partie réalisé par un expert. Du point de vue de la nature de l'apport sémantique, les classes sémantiques construites par des experts reflètent une connaissance partagée, elles sont une expression de la *doxa* alors que les dispositifs LUCIA reflètent un point de vue éventuellement partagé au sein d'un groupe social déterminé. Leur validité peut être remise en cause par d'autres individus que ceux qui les ont construits. Dans ce sens, un système centré sur l'utilisateur peut aussi avoir comme intérêt d'être le support d'un travail collaboratif. Plus que centré sur l'homme, le modèle est centré sur l'individu et le groupe d'individus. Sa malléabilité et sa dynamique, conférées par les principes d'une *sémantique légère* pour le TAL, lui allouent une valeur finalement négociable qui nous paraît être une nouvelle voie en informatique particulièrement adaptée aux attentes actuelles des utilisateurs.

¹³⁸ Nous ne reprenons pas à notre compte les propos initiaux de Pincemin qui parlent également d'une économie de *moyens*. Si par exemple, cette économie peut effectivement être, par rapport à une expertise humaine, de l'ordre du temps passé à l'exercice, les moyens mis en œuvre ne sont pas nécessairement « économiques » – les premiers chapitres de cette thèse l'ont montré.