



UNIVERSITE DE CAEN / BASSE-NORMANDIE

**UFR des Sciences
École Doctorale SIMEM**

THÈSE

présentée et soutenue publiquement le 7 décembre 2004
pour l'obtention du

Doctorat de l'Université de Caen / Basse-Normandie
Spécialité Informatique

par Vincent PERLERIN

Sémantique légère pour le document

*Assistance personnalisée pour l'accès au document
et l'exploration de son contenu*

Membres du jury

- M. Benoît Habert Professeur de l'Université de Paris X, Nanterre (*rapporteur*)
M. Ioannis Kanellos Professeur de l'ENST de Brest (*rapporteur*)
Mme. Anne Nicolle Professeur de l'Université de Caen (*directrice de thèse*)
M. François Rastier Directeur de recherche au CNRS
Mme. Pascale Sébillot Maître de conférences HDR de l'Université de Rennes I
M. Pierre Beust Maître de conférences de l'Université de Caen

More than machinery, we need humanity.
C. Chaplin, The Great Dictator, discours final

Remerciements

Je tiens à remercier l'ensemble des personnes qui m'ont aidé et soutenu dans mon travail. En particulier, j'adresse ma gratitude...

À messieurs Benoît Habert et Ioannis Kanellos pour m'avoir fait l'honneur de rapporter cette thèse et m'avoir prodigué de nombreux conseils pour améliorer ce document. (Meta;tau'ta oukēvi se ep̄w nikah).

À madame Pascale Sébillot et monsieur François Rastier pour leur participation à la commission d'examen de ces travaux.

À madame Anne Nicolle pour la confiance et la liberté qu'elle m'a accordées durant ce travail et pour sa très grande disponibilité lors de la rédaction de ce document.

À Pierre Beust et Stéphane Ferrari : leur soutien indéfectible, leurs conseils et leur humour m'ont été d'un grand secours. Ces travaux leur doivent beaucoup. Je leur suis d'autant plus reconnaissant que nos collaborations ont souvent mis au jour des points de vue différents. Leur soif d'invention et d'expérimentation fut plus forte que nos dissensions.

À Pierre-Sylvain Luquet, François Gérard et Jean-Pierre Leficus pour avoir supporté mes humeurs dans le bureau 164 et pour avoir organisé quelques respirations collectives indispensables.

À Charlotte Rault pour son amitié et son travail de relecture inespéré et méticuleux.

À Raymond Prunier, son travail titanesque et minutieux m'a été d'une grande utilité. Les nombreux ouvrages qu'il m'a donnés à lire lors de ces années de travail et les discussions que nous avons pu mener ensemble m'ont été d'un grand secours. À Patricia Prunier, pour ses rereadings et bien au-delà.

À mes parents parce-qu'ils sont bien sûr pour beaucoup dans tout cela.

À Blaise et Boubou pour leur soutien jusqu'au cœur de la tourmente. Merci de m'avoir appris qu'il fallait parfois perdre son temps. (*Bonne chance, monsieur Gorsky !*, Neil Amstrong). À Jocelyn Rey pour les mêmes raisons. À Aurore Bosquet pour d'autres raisons aussi.

À Christophe et Frédérique Turbout-Loew pour leur gentillesse et leur soutien. Aux autres membres d'APIDoC : l'association fut la meilleure école pour apprendre ce qu'était la recherche au-delà de ma discipline.

Aux membres du laboratoire GREYC avec qui mes rapports furent aussi divers qu'enrichissants. Je remercie particulièrement Patrice Enjalbert qui, par deux fois, a su comprendre que..., Antoine Widlöcher qui est un bien charmant compagnon de voyage et les membres de l'équipe ISLAnD : Jacques Vergne, Nadine Lucas et Bernard Morand.

À Francis Rousseau et Michel Bazin qui m'ont donné le goût de poursuivre dans cette voie.

À mes familles axonaises, amiénoises, rémoises et maintenant normandes sans qui tout cela ne serait pas même concevable.

Au népalais d'avoir su se taire pendant trois ans : et bien chantez maintenant !

Et enfin, à Anne-Laure pour avoir supporté tout cela pendant qu'elle travaillait stoïque, attentive et courageuse, à notre bonheur et à celui d'enfants en danger de mort.

à Anne-Laure.

Table des matières

Chapitre 1 Vers une <i>sémantique légère</i> pour le TAL	11
1.1 Un modèle de ressources sémantiques personnalisées.....	12
1.2 Définition des objectifs	15
1.2.1 Accès aux documents	15
1.2.2 Accès au contenu des documents	19
1.3 Démarche	22
1.3.1 Décrire des significations	22
1.3.2 Organiser les descriptions	25
1.3.3 L'interaction comme alternative aux approches classiques.....	27
1.4 Sémantique légère.....	28
1.4.1 Les ressources	28
1.4.2 Les processus.....	29
1.4.3 Vers une <i>sémantique légère</i> pour le TAL.....	30
1.5 Plan de la thèse	31
Chapitre 2 Fondements de l'approche	35
2.1 Accès aux documents et à leur contenu	36
2.1.1 Les ontologies et le web sémantique	36
2.1.2 Représentation des connaissances	40
2.1.3 Terminologie et linguistique.....	42
2.1.4 Subjectivité, hypermédias et interprétation	44
2.1.5 Conclusion.....	47
2.2 Fondements	51
2.2.1 Valeur saussurienne.....	52
2.2.2 Approche interprétative, Sémantique Interprétative	59
2.2.3 Modèle de catégorisation différentielle et modèle oppositionnel du sème.....	63
2.2.4 Interaction dans le système.....	65
2.3 Conclusion.....	67
Chapitre 3 Le modèle LUCIA	69
3.1 Entrées lexicales du système	71
3.1.1 Lexique de mots <i>versus</i> lexique de morphèmes	72
3.1.2 Détermination des entités	73
3.1.3 Critères de sélection	75
3.2 Modèle de catégorisation et de description	77
3.2.1 Les attributs	77
3.2.2 Les tables, les topiques.....	83
3.2.3 Les dispositifs.....	91
3.3 LUCIA, un modèle de représentation des mots dans le discours	96
3.3.1 Point de vue et jugement de l'utilisateur	97
3.3.2 Les mots dans le discours interprété.....	100
3.3.3 Approche complémentaire entre onomasiologie et sémasiologie.....	103

3.4	Expérience.....	104
3.5	Conclusion.....	110
Chapitre 4 Acquisition et structuration des ressources..... 111		
4.1	Les tâches	113
4.1.1	Aspects génériques des tâches.....	113
4.1.2	Veille documentaire	114
4.1.3	Étude d'une métaphore conceptuelle.....	120
4.1.4	Autres tâches	123
4.2	Corpus d'observation et acquisition	124
4.2.1	Définition du corpus d'observation.....	125
4.2.2	Extraction supervisée	127
4.2.3	MEMLABOR – Logiciel d'aide à l'acquisition.....	129
4.2.4	Première évaluation du lexique : THEMEEDITOR.....	141
4.2.5	Variantes morphosyntaxiques	147
4.3	LUCIABuilder – Logiciel interactif pour la construction de dispositifs.....	151
4.4	Propriétés des dispositifs.....	160
4.4.1	Exemple de dispositif.....	160
4.4.2	Symétrie du processus : des attributs aux dispositifs.....	163
4.5	Protocole de construction d'un dispositif	174
4.6	Conclusion.....	178
Chapitre 5 Analyses et interactions 181		
5.1	Projections des informations lexicales	182
5.2	Visualisation et interaction	190
5.2.1	Techniques de visualisation interactive	190
5.2.2	Interactions génériques et spécifiques	196
5.2.3	Facteurs à prendre en considération	207
5.3	Étude de la métaphore	210
5.3.1	Première expérience	211
5.3.2	Observations et résultats.....	213
5.3.3	Conclusions et perspectives pour l'étude de la métaphore	223
5.4	Veille documentaire.....	226
5.4.1	LUCIASearch.....	227
5.4.2	Exemple d'utilisation	232
5.4.3	Conclusions et perspectives sur le projet de veille documentaire.....	247
5.5	Évaluation	248
Conclusion.....		251
Bibliographie.....		255
Index		269
Glossaire.....		271