

Table des matières

Introduction	11
Chapitre 1. La sphère du lexique et des connaissances	13
1.1. Sémantique lexicale	13
1.1.1. Extension du sens lexical	14
1.1.1.1. Dénotation et connotation	14
1.1.1.2. La métaphore	15
1.1.1.3. La métonymie	17
1.1.1.4. La synecdoque	18
1.1.2. Relations paradigmatiques de sens	18
1.1.2.1. Champ sémantique et champ lexical	18
1.1.2.2. Les hiérarchies taxonomiques	20
1.1.2.3. Les hiérarchies méronymiques	21
1.1.2.4. Homonymie et polysémie	22
1.1.2.5. Synonymie	24
1.1.2.6. Opposition	26
1.1.2.7. Paronymie	28
1.1.2.8. Troponymie	28
1.1.3. Théories du sens lexical	28
1.1.3.1. L'approche aristotélicienne	28
1.1.3.2. Approche sémique ou componentielle	30
1.1.3.3. La sémantique du prototype	31
1.1.3.4. La théorie du lexique génératif	34
1.2. Bases de données lexicales	36
1.2.1. Standards pour l'encodage et l'échange des données	38
1.2.2. Encodages standard des caractères	38

1.2.2.1. SGML	39
1.2.2.2. XML	43
1.2.2.3. RDF	45
1.2.3. Standards pour le contenu	46
1.2.3.1. Le standard TEI	46
1.2.3.2. Le projet SALT	49
1.2.3.3. Le standard LMF	51
1.2.4. Systèmes d'écriture	54
1.2.4.1. Papillon	54
1.2.4.2. DEB	57
1.2.5. Quelques bases de données lexicales	59
1.2.5.1. WordNet	59
1.2.5.2. La base de noms propres Prolex	61
1.2.5.3. La base de données lexicales Brulex	63
1.2.5.4. Lexique	63
1.3. Représentation des connaissances et ontologies	64
1.3.1. Représentation des connaissances	64
1.3.1.1. Les formalismes de représentation des connaissances	65
1.3.1.2. Les réseaux sémantiques	66
1.3.1.3. Les graphes conceptuels	68
1.3.1.4. Les schémas	72
1.3.1.5. Les scripts	76
1.3.1.6. UNL	76
1.3.2. Ontologies	78
1.3.2.1. Méthodologie de développement d'ontologies	81
1.3.2.2. Structure d'une ontologie	82
1.3.2.3. Outils pour le développement d'ontologies	84
1.3.2.4. Quelques ontologies	84
Chapitre 2. La sphère de la sémantique	93
2.1. La sémantique combinatoire	93
2.1.1. La sémantique interprétative	93
2.1.2. La sémantique générative	98
2.1.3. La grammaire de cas	100
2.1.4. La sémantique interprétative de Rastier	103
2.1.5. La théorie sens-texte	110
2.2. La sémantique formelle	113
2.2.1. La logique propositionnelle	113
2.2.1.1. La notion de proposition	113
2.2.1.2. Les connecteurs logiques	114

2.2.1.3. Les formules bien formées	118
2.2.1.4. Les règles d'inférence	121
2.2.2. La logique de premier ordre	123
2.2.2.1. Termes	123
2.2.2.2. Prédicats	123
2.2.2.3. Quantificateurs	124
2.2.2.4. Formules bien formées (<i>fbf</i>)	126
2.2.2.5. Interprétation sémantique	128
2.2.3. Le lambda-calcul	130
2.2.3.1. La syntaxe des lambda-expressions	130
2.2.3.2. Types	132
2.2.3.3. La sémantique des lambda-expressions.	133
2.2.3.4. Les conversions.	133
2.2.3.5. Emploi du lambda-calcul dans l'analyse du langage naturel	134
2.2.4. D'autres types de logique	137

Chapitre 3. La sphère du discours et du texte 139

3.1. Analyse de discours et pragmatique	139
3.1.1. Notions fondamentales.	139
3.1.1.1. Discours <i>versus</i> parole.	140
3.1.1.2. Discours <i>versus</i> phrase	140
3.1.1.3. Discours <i>versus</i> récit.	140
3.1.1.4. Discours <i>versus</i> texte	140
3.1.2. Énonciation	141
3.1.2.1. Les déictiques (embrayeurs)	142
3.1.2.2. Les acteurs de la communication	142
3.1.2.3. Les circonstants de temps et de lieu.	144
3.1.3. Contexte, cotexte et intertextualité.	144
3.1.4. Structuration de l'information dans le discours	146
3.1.4.1. Le thème et le rhème.	146
3.1.4.2. La progression thématique	147
3.1.4.3. L'implicite, le présupposé et le sous-entendu	150
3.1.4.4. Donné <i>versus</i> nouveau.	152
3.1.5. Cohérence	153
3.1.6. Cohésion	155
3.1.6.1. Cohésion lexicale.	156
3.1.6.2. Anaphore	156
3.1.7. Ellipses	159
3.1.8. Séquences textuelles	160

3.1.9. Actes de langage	161
3.1.9.1. Les actes locutoires	162
3.1.9.2. Les actes illocutoires.	162
3.1.9.3. Les actes perlocutoires.	163
3.2. Approches computationnelles du discours.	163
3.2.1. Segmentation linéaire du discours	164
3.2.2. Théorie de la structure rhétorique et analyse automatique du discours.	165
3.2.3. Interprétation du discours : la TRD	171
3.2.4. Traitement des anaphores	176
3.2.4.1. L'approche syntaxique naïve de Hobbs.	177
3.2.4.2. L'approche discursive : la théorie du centrage (TC)	178
3.2.4.3. Les approches à base d'apprentissage.	183

Chapitre 4. La sphère des applications 187

4.1. Génie logiciel pour les linguiciels	187
4.1.1. Cycle de vie d'un linguiciel	187
4.1.2. Architectures logicielles pour le TAL	188
4.1.3. Architectures sérielles	189
4.1.4. Architectures centrées sur les données	191
4.1.5. Architectures orientées objet	195
4.1.6. Architectures multiagents	196
4.1.7. Coopération syntaxe-sémantique : des modèles cognitifs aux architectures logicielles.	198
4.1.8. Langages de programmation pour le TAL	204
4.1.9. Évaluation des systèmes de TAL.	206
4.1.9.1. Test structural	207
4.1.9.2. Évaluation quantitative	208
4.1.9.3. Évaluation qualitative	210
4.2. Traduction automatique (TA)	212
4.2.1. Pourquoi la traduction est-elle difficile ?	213
4.2.1.1. Divergences linguistiques	213
4.2.1.2. Ambiguïtés dans le texte de départ	214
4.2.1.3. Erreurs dans le texte de départ	214
4.2.2. Historique des systèmes de TA	215
4.2.2.1. Les débuts : fin des années 1940-1965	215
4.2.2.2. La deuxième génération (1965-1975).	216
4.2.2.3. Le regain d'intérêt (1975-1990)	216
4.2.2.4. La période récente	217
4.2.3. Typologie des systèmes de TA	217
4.2.4. Usage de la TA	219

4.2.5. Techniques de TA.	221
4.2.5.1. Approche directe	221
4.2.5.2. Approche par transfert	222
4.2.5.3. L'approche dite par pivot ou interlangue	225
4.2.5.4. La traduction à base d'exemples TBE	227
4.2.5.5. L'approche statistique	228
4.2.6. Exemple d'un système de traduction : Verbmobil	229
4.3. Recherche d'information (RI)	233
4.3.1. RI et les domaines connexes	233
4.3.2. L'information lexicale et la RI	235
4.3.2.1. La loi de Zipf	235
4.3.2.2. Filtrage des mots vides	238
4.3.2.3. Pondération des mots	238
4.3.2.4. La lemmatisation	240
4.3.3. Approches de recherches d'information.	241
4.3.3.1. Le balayage (scan) du fichier complet	241
4.3.3.2. Les listes inverses	241
4.3.3.3. Les fichiers de signature.	241
4.3.3.4. Les approches basées sur les groupements (<i>clustering</i>)	242
4.3.3.5. Les approches basées sur la similarité	244
4.3.3.6. Les approches itératives	248
4.3.3.7. Les approches bayésiennes	249
4.3.3.8. Les approches vectorielles : l'analyse sémantique latente (LSA)	250
4.4. Grandes masses de données (GMD) et extraction d'informations.	256
4.4.1. Données structurées, semi-structurées et non structurées.	256
4.4.2. Architectures des systèmes de traitement de GMD	257
4.4.3. Rôle du TAL dans le traitement des GMD	259
4.4.4. Extraction d'informations	260
4.4.4.1. Extraction d'entités nommées (EN)	262
4.4.4.2. Extraction de relations	265
4.4.4.3. Extraction d'événements	269
4.4.4.4. Analyse des sentiments et fouille d'opinions	271
Conclusion.	281
Bibliographie	285
Index	321
Sommaire de <i>Traitement automatique des langues et linguistique informatique 1</i>	325