

Table des matières

Introduction	9
Chapitre 1. La sphère des ressources linguistiques	15
1.1. La notion de corpus	15
1.2. Taxonomie des corpus	18
1.2.1. Ecrits <i>versus</i> parlés	18
1.2.2. Le point de vue historique	19
1.2.3. La langue des corpus	19
1.2.4. La représentativité thématique	21
1.2.5. Tranches d'âge des locuteurs	22
1.3. Qui collecte et distribue des corpus ?	22
1.3.1. Le projet Gutenberg	22
1.3.2. Le <i>Linguistic Data Consortium</i> (LDC)	22
1.3.3. <i>European Language Resource Agency</i> (ELRA)	23
1.3.4. <i>Open Language Archives Community</i> (OLAC)	23
1.3.5. Divers	23
1.4. Cycle de vie d'un corpus	24
1.4.1. Analyse des besoins	25
1.4.2. Conception de scénarii pour collecter le corpus	25
1.4.3. Collecte du corpus	26
1.4.4. Transcription.	29
1.4.5. Annotation du corpus.	31
1.4.6. Documentation du corpus	35
1.4.7. Analyse statistique des données	35
1.4.8. Usage des corpus dans le domaine de TAL.	36

1.5. Exemples de corpus existants	36
1.5.1. <i>American National Corpus</i> (ANC).	36
1.5.2. <i>Oxford English Corpus</i> (OEC)	36
1.5.3. Le corpus de l'office de tourisme de Grenoble.	37
Chapitre 2. La sphère de la parole	39
2.1. Etudes linguistiques de la parole	39
2.1.1. La phonétique	39
2.1.1.1. Phonétique articulatoire	41
2.1.1.2. Phonétique acoustique.	45
2.1.1.3. Phonétique auditive	50
2.1.1.4. Le système phonétique du français	52
2.1.2. La phonologie	59
2.1.2.1. Les notions de phonème et d'allophone	59
2.1.2.2. Les traits distinctifs	60
2.1.2.3. Les règles phonologiques	61
2.1.2.4. La syllabe	64
2.1.2.5. La phonologie autosegmentale	66
2.1.2.6. La théorie de l'optimalité (TO)	68
2.1.2.7. Prosodie et phonosyntaxe	70
2.2. Traitement automatique de la parole	72
2.2.1. Reconnaissance automatique de la parole.	73
2.2.1.1. Modèles acoustiques à base de HMM.	75
2.2.1.2. L'algorithme de Viterbi	82
2.2.1.3. Modèles de langage à base de n-grammes	86
2.2.2. Synthèse de la parole	89
2.2.2.1. Le traitement avant.	90
2.2.2.2. La synthèse concatenative	93
2.2.2.3. La synthèse par formants	95
2.2.2.4. La synthèse articulatoire	97
Chapitre 3. La sphère des mots	99
3.1. Eléments de morphologie	99
3.1.1. Typologie morphologique	100
3.1.2. Morphologie du français.	101
3.1.2.1. Types de morphème en français	101
3.1.2.2. Les allomorphes	104
3.1.2.3. Opérations de combinaison des morphèmes	104
3.1.2.4. L'abréviation	106

3.1.3. Les parties du discours	106
3.1.4. Les termes, collocations et colligations	110
3.2. Analyse morphologique automatique et étiquetage en parties du discours	111
3.2.1. La lemmatisation	111
3.2.1.1. La variation du successeur	112
3.2.1.2. L'algorithme de Porter.	113
3.2.1.3. Le groupement à base de n-grammes	114
3.2.2. Les expressions régulières pour l'analyse morphologique	115
3.2.3. Introduction informelle aux automates à états finis	118
3.2.4. La morphologie à deux niveaux et TFE.	122
3.2.5. L'étiquetage en parties du discours (<i>POS tagging</i>)	129
3.2.5.1. Approches statistiques	132
3.2.5.2. Approche à base de transformations.	134
Chapitre 4. La sphère de la syntaxe.	137
4.1. Notions syntaxiques de base	137
4.1.1. Délimitation du champ de la syntaxe	137
4.1.2. La notion de grammaticalité.	138
4.1.3. Les constituants syntaxiques	139
4.1.3.1. Les mots	140
4.1.3.2. Les propositions	142
4.1.3.3. Le syntagme.	142
4.1.3.4. Les <i>chunks</i>	146
4.1.3.5. La construction	148
4.1.3.6. La phrase.	148
4.1.4. Typologie syntaxique de la topologie et de l'accord	149
4.1.5. L'ambiguïté syntaxique	150
4.1.6. Spécificités syntaxiques du langage oral spontané	151
4.1.6.1. La topologie en français parlé	152
4.1.6.2. L'accord en genre et en nombre	152
4.1.6.3. Exemples de constructions syntaxiques et leurs spécificités à l'oral	153
4.1.6.4. Les extragrammaticalités du langage oral	155
4.2. Eléments de syntaxe formelle	156
4.2.1. Arbres d'analyse et règles de réécriture	156
4.2.2. Les langages et les grammaires formelles.	164
4.2.3. La hiérarchie des langages (Chomsky-Schützenberger)	165

4.2.4. Structure de traits et unification	174
4.2.5. La grammaire DCG.	181
4.3. Formalismes syntaxiques.	183
4.3.1. X-Barre.	183
4.3.2. La grammaire syntagmatique guidée par les têtes (HPSG).	191
4.3.2.1. Principes fondamentaux.	191
4.3.2.2. Les ST	192
4.3.2.3. Les règles morphologiques	197
4.3.2.4. Les règles syntaxiques.	198
4.3.2.5. Les principes sémantiques	200
4.3.2.6. Exemple d'analyse d'une phrase simple	200
4.3.3. La grammaire d'arbres adjoints lexicalisés (LTAG)	205
4.3.3.1. Principes fondamentaux.	205
4.3.3.2. Les unités de base	206
4.3.3.3. Les opérations de composition des arbres	208
4.3.3.4. La composition sémantique et l'opération d'unification.	210
4.4. L'analyse syntaxique automatique	212
4.4.1. Les automates à états finis (AEF)	214
4.4.2. Les réseaux de transition rékursifs (RTR).	214
4.4.3. Approche descendante (<i>top-down</i>)	218
4.4.4. Approche ascendante (<i>bottom-up</i>)	222
4.4.5. L'approche mixte : le coin gauche	225
4.4.6. L'analyse tabulaire (<i>chart</i>)	230
4.4.7. Analyse syntaxique probabiliste	235
4.4.8. Les réseaux neuronaux.	243
4.4.9. Algorithmes d'analyse pour les grammaires à base d'unification.	247
4.4.10. Les approches pour l'analyse syntaxique robuste	248
4.4.10.1. L'analyse partielle par segments (<i>chunking</i>)	248
4.4.10.2. Les approches sélectives.	250
4.4.10.3. L'analyse syntaxique des extragrammaticalités du langage oral	251
4.4.11. Les algorithmes de génération.	252

Bibliographie	253
--------------------------------	------------

Index	277
------------------------	------------