

---

# Table des matières

---

<b>Remerciements</b> . . . . .	13
<b>Chapitre 1. Introduction : le projet</b> . . . . .	15
1.1. Caractériser un ensemble de taille infinie . . . . .	18
1.2. L'ordinateur et la linguistique . . . . .	19
1.3. Les niveaux de la formalisation . . . . .	20
1.4. Hors sujet . . . . .	21
1.4.1. Poésie, jeux de mots . . . . .	21
1.4.2. Stylistique et rhétorique . . . . .	22
1.4.3. Anaphores, résolution de coréférence, désambiguïsation sémantique. . . . .	23
1.4.4. Calculs extralinguistiques . . . . .	24
1.5. Structure de cet ouvrage . . . . .	24
1.6. Exercices . . . . .	25
1.7. Liens Internet. . . . .	26
<b>Chapitre 2. D'autres approches</b> . . . . .	27
2.1. Analyse statistique textuelle . . . . .	27
2.1.1. Reconnaissance de la parole. . . . .	27
2.1.2. Traduction automatique . . . . .	29
2.1.3. Etiquetage statistique de textes . . . . .	34
2.1.4. Problèmes avec les méthodes statistiques. . . . .	35
2.1.5. Analyse linguistique, plutôt que stochastique . . . . .	41
2.2. Linguistique formelle . . . . .	42
2.2.1. XFST : un formalisme adapté à la morphologie . . . . .	42
2.2.2. LFG : un formalisme adapté à la syntaxe . . . . .	43

2.2.3. De nombreux formalismes incompatibles. . . . .	46
2.2.4. Une plateforme spécialement conçue pour la linguistique descriptive : NooJ . . . . .	46
2.2.5. Une seule notation pour cet ouvrage. . . . .	48
2.3. Informatique linguistique. . . . .	48
2.4. Conclusion . . . . .	50
2.5. Exercices . . . . .	50
2.6. Liens Internet. . . . .	50
<b>PREMIÈRE PARTIE. LES UNITÉS LINGUISTIQUES . . . . .</b>	<b>53</b>
<b>Chapitre 3. Formaliser l’alphabet . . . . .</b>	<b>57</b>
3.1. Bits et octets . . . . .	57
3.2. Numériser les informations . . . . .	58
3.3. Représenter les entiers naturels . . . . .	59
3.3.1. La notation décimale . . . . .	59
3.3.2. La notation binaire . . . . .	60
3.3.3. La notation hexadécimale . . . . .	61
3.4. Coder les caractères . . . . .	61
3.4.1. Standardisation des codages. . . . .	63
3.4.2. Lettres accentuées et diacritiques. . . . .	65
3.4.3. Les ligatures . . . . .	65
3.4.4. Variantes typographiques . . . . .	66
3.4.5. Rétablir les voyelles et accents . . . . .	67
3.4.6. Transcription des lettres étrangères . . . . .	68
3.4.7. Multi-alphabets . . . . .	68
3.4.8. Abréviations symboliques . . . . .	70
3.4.9. Les codages ASCII étendus . . . . .	70
3.4.10. Le codage Unicode . . . . .	71
3.5. Les ordres lexicographiques . . . . .	77
3.5.1. L’ordre lexicographique français. . . . .	78
3.5.2. Chaque langue a ses ordres lexicographiques . . . . .	79
3.5.3. Institutions et codages . . . . .	81
3.6. Classification des caractères . . . . .	82
3.6.1. Les lettres. . . . .	82
3.6.2. Les espaces. . . . .	83
3.7. Conclusion . . . . .	83
3.8. Exercices . . . . .	84
3.9. Liens Internet. . . . .	84

<b>Chapitre 4. Délimiter le vocabulaire.</b> . . . . .	85
4.1. Vocabulaires multiples et synchronie . . . . .	85
4.1.1. Evolution du vocabulaire . . . . .	87
4.1.2. Recensement explicite ou utilisation de règles ? . . . . .	90
4.1.3. La dérivation est une opération morpho-lexico-syntactico-sémantique . . . . .	90
4.1.4. Les dérivations sont imprévisibles . . . . .	91
4.1.5. Atomicité des mots dérivés . . . . .	92
4.2. Les Unités linguistiques atomiques (ALU) . . . . .	94
4.2.1. Classification des ALU . . . . .	95
4.2.2. Quelques conséquences de la classification. . . . .	98
4.3. Unités atomiques ou analysables ? . . . . .	99
4.3.1. Mots simples <i>versus</i> formes analysables morphologiquement . . . . .	99
4.3.2. Mots composés <i>versus</i> séquences analysables de mots simples . . . . .	102
4.4. Conclusion . . . . .	111
4.5. Exercices . . . . .	112
4.6. Liens Internet. . . . .	112
<b>Chapitre 5. Réutiliser les ressources existantes</b> . . . . .	115
5.1. Réutiliser les dictionnaires éditoriaux ? . . . . .	115
5.2. Incomplétude des dictionnaires éditoriaux . . . . .	115
5.3. Peut-on extraire de ces dictionnaires la liste des noms humains d'une langue ? . . . . .	116
5.4. L'orthographe : une norme ? . . . . .	121
5.4.1. Inadéquation de l'orthographe . . . . .	122
5.4.2. L'argument étymologique . . . . .	124
5.4.3. Le français est avant tout une langue parlée . . . . .	124
5.4.4. Pas de nombre en français . . . . .	125
5.4.5. Les marqueurs morphologiques . . . . .	126
5.4.6. Les marques de conjugaison . . . . .	127
5.4.7. La réforme de 1990. . . . .	127
5.4.8. Traitement automatique de la variation orthographique . . . . .	130
5.5. Conclusion . . . . .	132
5.6. Exercices . . . . .	132
5.7. Liens Internet. . . . .	133

<b>Chapitre 6. Les dictionnaires électroniques</b> . . . . .	135
6.1. Le système des dictionnaires DELA . . . . .	139
6.1.1. DELA et INTEX . . . . .	141
6.1.2. Du DELA à NooJ . . . . .	142
6.2. Les dictionnaires de <i>Dubois</i> et <i>Dubois-Charlier</i> . . . . .	148
6.2.1. Le <i>Dictionnaire électronique des mots</i> . . . . .	149
6.2.2. Le dictionnaire <i>Les verbes français</i> . . . . .	150
6.2.3. Formalisation des dictionnaires DEM et LVF . . . . .	152
6.2.4. Le champ DERIVATION . . . . .	154
6.2.5. Le champ CONSTRUCTION . . . . .	155
6.2.6. Sélection distributionnelle . . . . .	159
6.2.7. Les dictionnaires DEM et LVF dans NooJ . . . . .	160
6.3. Conclusion . . . . .	161
6.4. Exercices . . . . .	162
6.5. Liens Internet. . . . .	162
<b>DEUXIÈME PARTIE. LANGAGES, GRAMMAIRES ET ANALYSEURS</b> . . . . .	165
<b>Chapitre 7. Langages, grammaires et machines</b> . . . . .	169
7.1. Définitions . . . . .	169
7.1.1. Lettres et alphabets . . . . .	169
7.1.2. Mots et langages . . . . .	170
7.1.3. ALU, vocabulaires, phrases et langages. . . . .	170
7.1.4. Le mot vide . . . . .	171
7.1.5. Le langage libre . . . . .	172
7.1.6. Grammaires . . . . .	172
7.1.7. Machines . . . . .	173
7.2. Grammaires génératives . . . . .	174
7.3. Hiérarchie de Chomsky-Schützenberger. . . . .	175
7.3.1. Les langues naturelles sont des ensembles récursivement énumérables . . . . .	178
7.3.2. Les formalismes linguistiques . . . . .	178
7.4. L'approche de NooJ. . . . .	180
7.4.1. Une approche multiple . . . . .	180
7.4.2. Une notation unifiée . . . . .	181
7.4.3. Architecture en cascade . . . . .	183
7.5. Conclusion . . . . .	183
7.6. Exercices . . . . .	184
7.7. Liens Internet. . . . .	185

<b>Chapitre 8. Grammaires et langages rationnels</b> . . . . .	187
8.1. Expressions rationnelles . . . . .	187
8.1.1. Quelques exemples de grammaires rationnelles . . . . .	191
8.2. Graphes à états finis . . . . .	193
8.3. Théorème de Kleene . . . . .	194
8.4. Expressions rationnelles à production et transducteurs finis . . . . .	199
8.5. Extensions des grammaires rationnelles . . . . .	201
8.5.1. Les symboles lexicaux . . . . .	202
8.5.2. Les symboles syntaxiques . . . . .	204
8.5.3. Symboles définis par des grammaires . . . . .	205
8.5.4. Les opérateurs spéciaux . . . . .	207
8.6. Conclusion . . . . .	210
8.7. Exercices . . . . .	211
8.8. Liens Internet. . . . .	211
<b>Chapitre 9. Grammaires hors contexte et langages algébriques</b> . . . . .	213
9.1. Récursivité . . . . .	215
9.1.1. Récursivité à droite . . . . .	217
9.1.2. Récursivité à gauche . . . . .	218
9.1.3. Récursivité au milieu . . . . .	219
9.2. Arbres de dérivation. . . . .	221
9.3. Conclusion . . . . .	224
9.4. Exercices . . . . .	225
9.5. Liens Internet. . . . .	225
<b>Chapitre 10. Grammaires et langages contextuels</b> . . . . .	227
10.1. L’approche de NooJ . . . . .	228
10.1.1. Le langage $a^n b^n c^n$ . . . . .	229
10.1.2. Le langage $a^{2^n}$ . . . . .	232
10.1.3. Traiter les réduplications . . . . .	233
10.1.4. Les accords grammaticaux . . . . .	234
10.1.5. Contraintes lexicales dans les grammaires morphologiques . . . . .	236
10.2. Les contraintes contextuelles de NooJ . . . . .	237
10.3. Les variables . . . . .	239
10.3.1. Calcul de la valeur d’une variable, héritage. . . . .	241
10.4. Conclusion . . . . .	242
10.5. Exercices . . . . .	243
10.6. Liens Internet . . . . .	243

<b>Chapitre 11. Grammaires et langages non restreints</b> . . . . .	245
11.1. Adéquation linguistique. . . . .	246
11.2. Conclusion . . . . .	249
11.3. Exercice . . . . .	249
11.4. Liens Internet . . . . .	250
<b>TROISIÈME PARTIE. ANALYSES LINGUISTIQUES AUTOMATIQUES</b> . . . . .	251
<b>Chapitre 12. La structure d’annotation du texte</b> . . . . .	255
12.1. Analyser un texte. . . . .	255
12.2. Annotations . . . . .	256
12.2.1. Limites de la représentation XML/TEI . . . . .	258
12.3. La structure d’annotation du texte (TAS) . . . . .	259
12.4. Projet NooJ . . . . .	262
12.5. Liens Internet . . . . .	263
<b>Chapitre 13. Analyse lexicale</b> . . . . .	265
13.1. Analyse typographique . . . . .	265
13.1.1. Analyse des lettres et des séparateurs . . . . .	266
13.1.2. Apostrophe et trait d’union . . . . .	267
13.1.3. Analyser les nombres . . . . .	270
13.1.4. Segmenter un texte en phrases . . . . .	271
13.2. Reconnaissance des mots simples . . . . .	273
13.2.1. Contractions . . . . .	275
13.2.2. Contractions avec <i>à</i> . . . . .	275
13.2.3. Contractions en <i>du</i> . . . . .	276
13.2.4. Contractions en <i>de</i> . . . . .	277
13.2.5. Contractions en <i>des</i> . . . . .	278
13.2.6. Contractions avec <i>dit</i> . . . . .	279
13.2.7. La casse . . . . .	279
13.2.8. Accents ou voyelles manquants . . . . .	281
13.2.9. Les noms propres . . . . .	283
13.2.10. Les abréviations . . . . .	285
13.2.11. Variation orthographique . . . . .	285
13.2.12. Les chiffres romains . . . . .	286
13.2.13. L’élision en français . . . . .	287

13.3. Analyses morphologiques . . . . .	292
13.3.1. La morphologie flexionnelle . . . . .	293
13.3.2. Flexion des mots composés . . . . .	300
13.3.3. Morphologie dérivationnelle . . . . .	301
13.3.4. Combinaison dérivation/flexion . . . . .	303
13.3.5. Morphologie lexicale . . . . .	305
13.3.6. Dérivations productives . . . . .	306
13.3.7. Agglutinations . . . . .	307
13.4. Reconnaissance des mots composés . . . . .	312
13.4.1. Motivations . . . . .	313
13.4.2. Variations . . . . .	317
13.4.3. Variation sur la casse . . . . .	317
13.4.4. Variation des séparateurs . . . . .	318
13.4.5. Variation en nombre . . . . .	318
13.4.6. Variations idiosyncratiques . . . . .	319
13.4.7. Familles terminologiques . . . . .	319
13.4.8. Ambiguïtés . . . . .	321
13.4.9. Formes explicites et sous-structures . . . . .	324
13.5. Reconnaissance des expressions discontinues . . . . .	326
13.5.1. Constituant caractéristique . . . . .	328
13.5.2. Variation du constituant caractéristique . . . . .	330
13.5.3. Variation du verbe support . . . . .	331
13.5.4. Levée d'ambiguïté . . . . .	332
13.5.5. Annotations discontinues . . . . .	337
13.6. Conclusion . . . . .	339
13.7. Projet NooJ . . . . .	340
<b>Chapitre 14. Analyses syntaxiques locales . . . . .</b>	<b>341</b>
14.1. Variantes terminologiques . . . . .	341
14.2. Contextes des mots grammaticaux . . . . .	341
14.3. Entités nommées . . . . .	342
14.4. Levée d'ambiguïté automatique . . . . .	345
14.4.1. Grammaire des pronoms personnels . . . . .	348
14.4.2. Grammaire des pronoms postposés . . . . .	349
14.4.3. Levée d'ambiguïté des déterminants/pronoms <i>le, la, les, l'</i> . . . . .	352
14.5. Conclusion . . . . .	353
14.6. Projet NooJ . . . . .	354
14.7. Liens Internet . . . . .	354

<b>Chapitre 15. Analyse syntaxique structurelle</b> . . . . .	355
15.1. Unités linguistiques atomiques complexes. . . . .	355
15.2. Annotations structurées . . . . .	360
15.3. Les ambiguïtés . . . . .	363
15.4. Arbres syntaxiques et arbres de dérivation. . . . .	366
15.5. Arbre de dépendance sémantique . . . . .	369
15.6. Levée d’ambiguïté automatique et transparente. . . . .	371
15.7. Conclusion . . . . .	374
15.8. Projet NooJ . . . . .	375
<b>Chapitre 16. Analyses transformationnelles</b> . . . . .	377
16.1. Implémentation des transformations . . . . .	380
16.2. Problèmes théoriques . . . . .	385
16.2.1. Nombre d’opérateurs transformationnels à implémenter . . . . .	385
16.2.2. Ordre des transformations . . . . .	388
16.2.3. Equivalence des chaînes de transformations . . . . .	388
16.2.4. Ambiguïté sémantique . . . . .	389
16.2.5. Phrases théoriques. . . . .	390
16.2.6. Formaliser l’ensemble des chaînes de transformations ? . . . . .	391
16.3. L’analyse transformationnelle avec NooJ . . . . .	392
16.3.1. Utiliser les grammaires en mode « génération » . . . . .	393
16.3.2. Les arguments de la transformation . . . . .	394
16.4. Réponses aux questions. . . . .	397
16.5. Analyse sémantique . . . . .	398
16.6. Traduction automatique. . . . .	400
16.7. Conclusion . . . . .	404
16.8. Projets NooJ. . . . .	404
<b>Chapitre 17. Conclusion</b> . . . . .	407
<b>Bibliographie</b> . . . . .	411
<b>Index</b> . . . . .	421