

Table des matières

Table des figures.....	XI
Remerciements	XIII
0. Introduction	1
0.1. Objectif de l'étude	2
0.2. Plan de l'étude.....	3
1. État de la question	5
1.1. L'intérêt d'un FEW informatisé	5
1.1.1. L'accessibilité des données	5
1.1.2. Les consultations transversales	6
1.1.3. L'exhaustivité des résultats.....	8
1.1.4. La mise à jour du dictionnaire	10
1.1.5. Appels à l'informatisation du FEW	11
1.2. L'index sélectif des formes du FEW	12
1.2.1. L'intérêt d'un index du FEW	12
1.2.2. Le choix des informations à indexer	13
1.2.3. Les étapes de la réalisation	15
1.2.4. Les problèmes posés par l'indexation du FEW	17
1.3. Le prototype de FEWi sur Access	19
1.3.1. Synthèse	19
1.3.2. Analyse critique	27
1.3.3. Enseignements à tirer de ce travail.....	29
1.4. Le balisage des articles en cours de rédaction	30
1.4.1. Contexte historique.....	30
1.4.2. Différences structurelles avec la rétroconversion du FEW	31
1.4.3. Analyse du balisage	32
1.4.4. Conclusion	42

1.5. Les autres projets informatiques autour du FEW	43
1.5.1. Saisie du fichier onomasiologique	44
1.5.2. Informatisation des étymologies doubles	44
1.5.3. Version électronique de l'index sélectif des formes	44
1.5.4. Version électronique du <i>Beiheft</i>	45
1.5.5. Mise en ligne du manuel d'utilisation du FEW	45
1.5.6. Cartographie informatique.....	45
1.6. Conclusion	45
2. Le FEW vu par l'utilisateur	47
2.1. Introduction	47
2.2. Les utilisateurs du FEW.....	48
2.2.1. Utilité du FEW.....	48
2.2.2. Domaines d'utilisation et catégories d'utilisateurs.....	49
2.3. Itinéraires d'utilisation actuels	51
2.3.1. Comportements de consultation	51
2.3.2. Comportements de lecture	53
2.3.3. Conclusion	57
2.4. Itinéraires d'utilisation souhaités	58
2.4.1. Souhaits en relation avec la consultation du FEW.....	58
2.4.2. Souhaits en relation avec la lecture du FEW	62
2.4.3. Souhaits en relation avec la mise à jour du FEW.....	69
2.4.4. Conclusion	73
2.5. Deux visions du FEW	74
2.5.1. Des divergences apparentes	74
2.5.2. Complémentarité des deux dimensions.....	76
2.6. L'implicite fewien	77
2.6.1. Relevé de l'implicite.....	77
2.6.2. Analyse structurelle	78
2.6.3. Analyse du point de vue du décodage.....	79
2.7. Conclusion	83
3. Modélisation du discours fewien	85
3.1. Introduction	85
3.2. Unités de traitement.....	86

TABLE DES MATIÈRES

3.3. Principes de base	88
3.3.1. Redescente des informations.....	88
3.3.2. Limitations à la résolution de l'implicite	90
3.3.3. Limitation à l'identification des informations	92
3.4. Formalisation en XML	93
3.4.1. Choix du formalisme XML	93
3.4.2. Usage raisonné de la syntaxe XML	93
3.4.3. Formalisation des principes de base	94
3.5. Modélisation de l'étage supérieur : l'article	95
3.5.1. Redescente des informations macro- et superstructurelles	95
3.5.2. Traitement des matériaux d'origine inconnue	98
3.5.3. Le champ de l'entrée.....	98
3.5.4. Les renvois internes	102
3.6. Modélisation de l'étage inférieur : la cellule lexicale	103
3.6.1. Identification de la cellule lexicale	103
3.6.2. Structure de la cellule	103
3.6.3. Modélisation des molécules obligatoires	106
3.6.4. Modélisation des molécules facultatives.....	109
3.6.5. Formalisation XML de la cellule lexicale	117
3.7. Modélisation des étages intermédiaires.....	117
3.7.1. Séparation d'un article en différents champs	117
3.7.2. Structuration des éléments de documentation	118
3.7.3. Structuration des éléments de commentaire et de note	123
3.7.4. Mise en relation des notes et du commentaire avec la documenta- tion	127
3.7.5. Traitement de l'implicite microstructurel	129
3.8. Application du modèle sur un article du FEW	130
3.8.1. Résumé des types d'information identifiés	130
3.8.2. Exemple d'application du modèle	131
3.9. Conclusion	135
4. Architecture du système de rétroconversion.....	137
4.1. Introduction	137
4.2. Du papier au support numérique	138
4.2.1. Saisie par numérisation ou saisie manuelle ?.....	138

TABLE DES MATIÈRES

4.2.2. Codage des caractères	139
4.3. Du document numérique brut à un document XML avec balisage de base	142
4.3.1. Découpage en unités atomiques de rétroconversion	142
4.3.2. Introduction d'un balisage typographique	143
4.3.3. Validation des documents XML avec balisage de base	144
4.4. Du document XML avec balisage de base au document final	145
4.4.1. Architecture du système de rétroconversion	145
4.4.2. Représentation du texte traité par les algorithmes	147
4.4.3. Interaction du linguiste avec le système de rétroconversion	150
4.5. Conclusion	151
5. Algorithmes de rétroconversion	153
5.1. Introduction	153
5.2. Méthodologie de conception d'un algorithme de balisage	154
5.2.1. Première étape : identification des indicateurs de reconnaissance	155
5.2.2. Deuxième étape : traduction en termes informatiques	157
5.2.3. Troisième étape : écriture des algorithmes	163
5.2.4. Quatrième et cinquième étapes : validation et optimisation	164
5.3. Algorithmes de prétraitement	164
5.3.1. Algorithmes de détection d'erreurs de saisie ou de numérisation	165
5.3.2. Algorithmes de correction d'incohérences du FEW	168
5.3.3. Algorithme interprétant les tirets de fin de ligne	170
5.4. Algorithmes de balisage	174
5.4.1. Balisage des notes	174
5.4.2. Balisage des appels de note	179
5.4.3. Balisage de l'entrée	186
5.4.4. Balisage des étymons	188
5.4.5. Balisage des langues d'étymon	194
5.4.6. Balisage des renvois internes	198
5.4.7. Balisage des signatures	208
5.4.8. Balisage des définitions	216
5.4.9. Balisage des sigles bibliographiques	218
5.4.10. Balisage des sigles géolinguistiques	223
5.4.11. Séparation de la documentation et du commentaire	229

5.4.12. Balisage des marqueurs alphanumériques	235
5.4.13. Séparation de la documentation et du commentaire (2)	242
5.4.14. Balisage des affixes	246
5.4.15. Balisage des dates	248
5.4.16. Balisage des catégories grammaticales	251
5.4.17. Balisage des signifiants	253
5.4.18. Balisage des concepts	257
5.4.19. Balisage des précisions	260
5.4.20. Balisage des attestations	264
5.4.21. Balisage des titres	265
5.4.22. Balisage des cellules lexicales	269
5.4.23. Réunification des paragraphes	273
5.4.24. Structuration de la documentation	275
5.5. Algorithmes de post-traitement	279
5.5.1. Identification des parties non balisées	280
5.5.2. Identification des parties de cellules lexicales non balisées	281
5.5.3. Identification des parties balisées	281
5.5.4. Identification des balises occupant une place suspecte	282
5.5.5. Extraction des cellules lexicales	282
5.6. Séquençage des algorithmes	283
5.6.1. Dépendances entre algorithmes	283
5.6.2. Graphe des dépendances	284
5.6.3. Choix du séquençage	287
5.7. Conclusion	289
6. Analyse et exploitation des résultats	291
6.1. Introduction	291
6.2. Exemple d'article rétroconverti	292
6.2.1. Explicitation de la dimension monographique	292
6.2.2. Explicitation de la dimension thesaurus	294
6.2.3. Informations non reconnues	296
6.3. Rétroconversion du premier corpus de test	298
6.3.1. Temps de rétroconversion	299
6.3.2. Analyse du balisage inséré	299
6.3.3. Conclusion	304

TABLE DES MATIÈRES

6.4. Rétroconversion du second corpus de test.....	305
6.4.1. Temps de rétroconversion.....	306
6.4.2. Analyse du balisage inséré.....	306
6.4.3. Conclusion.....	307
6.5. Modalités d'exploitation des articles rétroconvertis.....	307
6.5.1. Informatisation du <i>Beiheft</i>	308
6.5.2. Modalités de consultation.....	310
6.5.3. Modalités de lecture.....	312
6.5.4. Modalités de mise à jour du FEW.....	313
6.6. Conclusion.....	314
7. Conclusion générale.....	317
8. Bibliographie.....	321
8.1. Liste des sigles bibliographiques.....	321
8.2. Ouvrages.....	322