Inhaltsverzeichnis

Einleitung 1

Aufgabestellung 3

Teil I: Anforderungen

1. Design-Philosophie: Das Lexikon als Datenbank 9
   1.1 Datenbank-Grundsätze und deren Relevanz für das DBS 10
   1.2 Das ANSI-SPARC 3-Schema-Modell 13
   1.3 Logische Datenmodelle 16

2. Klassifizierung des ET-Systems 19
   2.1 Benutzerbezogene Kriterien 19
      2.1.1 Verfügbarkeit 19
      2.1.2 Kompetenz 20
      2.1.3 Leistungsverteilung Mensch-Maschine 20
   2.2 Linguistische Kriterien 22
      2.2.1 Linguistische Theorien 22
      2.2.2 Sprachpaarbindung 23
      2.2.3 Übersetzungsebene 23
      2.2.4 Kontextberücksichtigung 25
   2.3 Informatische Kriterien 25
      2.3.1 Trennung von Daten und Programmen 25
      2.3.2 Programmiersprachen 26
   2.4 Generationen-Klassierung 27
      2.4.1 Erste Generation 27
      2.4.2 Zweite Generation 28
      2.4.3 Dritte Generation 29

3. Die ET-Software 31
   3.1 Design-Philosophie 31
   3.2 Das ET-Produktionssystem 32
      3.2.1 Die Datenbasis 32
      3.2.2 Die Regeln 40
      3.2.3 Die Interpretations-Prozesse 41
      3.2.4 Modularisierung 41
      3.2.5 Deklarativität und Prozeduralität 45
   3.3 Realisierung des Prototyping 48

4. Die ET-Linguistik 53
   4.1 Generelle Philosophie 53
   4.2 Grundlegende Terminologie 54
      4.2.1 Text, Wort-Einheit, Repräsentation 54
      4.2.2 Lexikale Einheit, lexikaler Transfer 55

4.3 Syntakto-semantische Ebene im Lexikon 56
4.4 Morphologische Ebene im Lexikon 56
4.4.1 Aufgabenbereich 56
4.4.2 Der 'ambitiöse' ET-Vorschlag 57
4.4.3 Probleme und Alternativen 60
4.4.3.1 LE-Definition 60
4.4.3.2 Linguistische Gesamttheorie 61
4.4.3.3 Morphologische Strategie, Morphologie-Sprachen 63
4.4.3.4 Semi-Regularität der Morphologie 63
4.4.3.5 Derivationen und Komposita mit selbständiger Bedeutung . . . 63
4.4.3.6 Lexikaler Transfer 64
4.4.3.7 Konsistenz und Redundanz 66
4.4.3.8 Arbeitsaufwand 69

5. Existierende maschinenlesbare Lexika 73
5.1 Anwendungsgebiete 73
5.1.1 Kommerzielle Anwendungen 73
5.1.2 Angewandte Forschung 76
5.1.3 Theoretische Forschung 77
5.2 Herkunft 79
5.3 Beispiele 80
5.3.1 Bearbeitung des 'Deutschen Wörterbuchs' 81
5.3.2 SADAW 84
5.3.3 SYSTRAN 87
5.4 Projekte zur Datenaufbereitung 91
5.4.1 Bonn 91
5.4.2 Kyoto 93
5.4.3 Cambridge 94
5.5 Schlussfolgerungen 94

Teil II: Prototyp-Entwurf

Einleitung 99

1. Ausgangsbasis: Koskenniemis 2-Ebenen-Modell 101
1.1 Generelle Eigenschaften 101
1.1.1 Sprachunabhängigkeit 101
1.1.2 'Mächtigkeit' und Problemorientierung 101
1.1.3 Anlehnung an linguistische Theorien 102
1.1.4 Bidirektionalität 102
1.1.5 Effiziente Verarbeitung 102
1.2 Das Modell 103
1.2.1 Das Alphabet 104
1.2.2 Die Grammatik 105
1.2.3 Die Datenbasis 112
1.2.4 Realisierung des Systems durch ein Programm 113

1.3 Schwachstellen 115
1.3.1 Formale Definition 116
1.3.2 Zeichensatz 116
1.3.3 Unabhängigkeit von Grammatik, Alphabet und Datenbasis 116
1.3.4 Trennung zwischen verschiedenen Schemata 117
1.3.5 Fazit 120

2. System-Übersicht 123
2.1 Die DBA- bzw. Linguisten-Schnittstelle 124
2.1.1 EMACS 125
2.1.2 YACC 125
2.1.3 LEX 125
2.1.4 Die Funktion der 'Werkzeuge' 126
2.2 Das interne Schema und die eigentlichen Daten 127
2.3 Die SLLP- und Lexikographen-Schnittstellen 129
2.4 Fazit 130

3. Die DDL 131
3.1 Die Abstraktions-Funktion von EMACS 132
3.1.1 Objekt-Orientierung 133
3.1.2 Das Zeichensatz-Problem 136
3.1.3 Abbildung auf eine YACC-konforme Syntax 138
3.2 Alphabet-Definition 140
3.2.1 Stellung im Gesamtsystem 140
3.2.2 Syntax-Definition 140
3.2.3 Erläuterungen und semantische Restriktionen 142
3.3 Typen-Definition 143
3.3.1 Stellung im Gesamtsystem 143
3.3.2 Syntax-Definition 144
3.3.3 Erläuterungen und semantische Restriktionen 145
3.4 Grammatik-Definition 146
3.4.1 Stellung im Gesamtsystem 146
3.4.2 Syntax-Definition 146
3.4.3 Erläuterungen und semantische Restriktionen 148
3.5 Datenbasis-Definition 149
3.5.1 Stellung im Gesamtsystem 149
3.5.2 Ausgangssituation 149
3.5.3 Das Redundanzproblem 151
3.5.4 Die reine Merkmalskategorisierung 152
3.5.4.1 Distinktive Fallunterscheidungen 152
3.5.4.1.1 Syntax-Definition 152
3.5.4.1.2 Erläuterungen und semantische Restriktionen 153
3.5.4.2 Generelle Fallunterscheidungen 156
3.5.4.2.1 Funktion 156
3.5.4.2.2 Syntax-Definition 157
3.5.4.2.3 Erläuterungen und semantische Restriktionen 158

3.5.5 Deklarationen 161
3.5.5.1 Funktion 161
3.5.5.2 Syntax-Definition 162
3.5.5.3 Erläuterungen und semantische Restriktionen 164
3.5.5.3.1 Merkmale des Mengen-Typs (Regeln 6-13) 164
3.5.5.3.2 Mengen von Merkmalen (Regeln 14-22) 165
3.5.5.3.3 Generelle Fallunterscheidungen (Regeln 23-36) 165
3.5.6 Direkte Assoziation von Merkmalen zu ME 168
3.5.6.1 LE-Klassen 168
3.5.6.1.1 Syntax-Definition 168
3.5.6.1.2 Erläuterungen und semantische Restriktionen 171
3.5.6.1.2.1 Grundlegendes Konzept 171
3.5.6.1.2.2 Fixierungen in Grundformen 174
3.5.6.1.2.3 Fixierungen wegen variierender Stammformen 175
3.5.6.1.2.4 Weitere Beispiele und Verarbeitung 176
3.5.6.1.2.5 Die implementierende Funktion 178
3.5.6.1.2.6 Das Konzept der Fortsetzungsklassen 179
3.5.6.2 Transfer-Einheiten 182
3.5.6.2.1 Syntax-Definition 182
3.5.6.2.2 Erläuterungen und semantische Restriktionen 183
3.5.6.3 LE-Einträge 186

4. Das interne Schema 187
4.1 Alphabet-Definition 188
4.2 Typen-Definition 188
4.3 Grammatik-Definition 190
4.4 Datenbasis-Definition 191
4.4.1 Distinktive Fallunterscheidungen 192
4.4.2 Generelle Fallunterscheidungen 193
4.4.3 Deklarationen 194
4.4.4 LE-Klassen 196
4.4.5 Einträge 198
4.4.6 LE-Einträge 199
4.5 Fazit 200

5. Monitor- und Manipulationsfunktionen der DBA-Schnittstelle 201
5.1 Ohne externe Einträge 202
5.1.1 Oberste Ebene (schema ... end schema) 202
5.1.2 Zweitoberste Ebene (z.B.: English ... end English) 202
5.1.3 Im Alphabet-Fenster 203
5.1.4 Im Typen-Fenster 204
5.1.5 Im Grammatik-Fenster 204
5.1.6 In Fenstern für distinktive Fallunterscheidungen 205

5.1.7 In Fenstern von generellen Fallunterscheidungen 206
5.1.8 In Fenstern von Deklarationen 206
5.1.9 In Fenstern von LE-Klassen 206
5.1.10 In Fenstern von Transfer-Einheiten 207
5.1.11 In Fenstern von LE-Einträgen 207
5.2 Mit externen Einträgen 207
5.2.1 Im Normalzustand 208
5.2.1.1 Oberste Ebene (schema ... end schema) 208
5.2.1.2 Zweitoberste Ebene (z.B.: English ... end English) 208
5.2.1.3 Im Alphabet-Fenster 209
5.2.1.4 Im Typen-Fenster 209
5.2.1.5 Im Grammatik-Fenster 209
5.2.1.6 In Fenstern für distinktive Fallunterscheidungen 210
5.2.1.7 In Fenstern von generellen Fallunterscheidungen 212
5.2.1.8 In Fenstern von Deklarationen 212
5.2.1.9 In Fenstern von LE-Klassen 212
5.2.1.10 In Fenstern von Transfer-Einheiten 213
5.2.1.11 In Fenstern von LE-Einträgen 213
5.2.2 Im Übergangszustand 213

6. Evaluation 215
6.1 Die SLLP-Schnittstelle 215
6.1.1 Vorschlag zur Übersetzungs-Strategie 215
6.1.1.1 Analyse 215
6.1.1.2 Transfer 217
6.1.1.3 Synthese 217
6.1.2 Zusammenfassung der sieben Schnittstellen-Funktionen 218
6.2 Die Lexikographen-Schnittstelle 221
6.2.1 Das Monitorprogramm 'browse' 221
6.2.2 Das Manipulationsprogramm 'manipulate' 222
6.3 Erweiterungsmöglichkeiten 224
6.3.1 Schnittstellen 224
6.3.2 Integration ins Übersetzungsprogramm 225
6.3.3 In der DDL 225
6.3.3.1 Alphabet-Definition 225
6.3.3.2 Typen-Definition 226
6.3.3.3 Grammatik-Definition 226
6.3.3.4 Datenbasis-Definition 226
6.3.3.4.1 Distinktive Fallunterscheidungen 226
6.3.3.4.2 Generelle Fallunterscheidungen 226
6.3.3.4.3 Deklarationen 227
6.3.3.4.4 LE-Klassen 227
6.3.3.4.5 Transfer-Einheiten 227
6.3.3.4.6 LE-Einträge 228
6.4 Fazit 228

Appendix: Liste der Forderungen 231

Bibliographie 235