

Contents

Overview	1
1. Introduction	3
1.1 Incremental Natural Language Processing	3
1.2 Incremental Speech Understanding	11
1.3 Incremental Architectures and the Architecture of MILC	15
1.4 Summary	24
2. Graph Theory and Natural Language Processing	25
2.1 General Definitions	25
2.2 The Use of Word Graphs for Natural Language Processing Systems	30
2.3 Evaluation of Word Graphs: Size and Quality Measures	34
2.4 Evaluation of Word Graphs: Quality Measures	44
2.5 Further Operations on Word Graphs	48
2.5.1 Removing Isolated Silence	48
2.5.2 Removing Consecutive Silence	49
2.5.3 Removing All Silence Edges	51
2.5.4 Merging Mutually Unreachable Vertices	51
2.6 Hypergraphs	52
2.6.1 Formal Definition of Hypergraphs	54
2.6.2 Merging of Hyperedges	56
2.6.3 Combination of Hyperedges	59
2.7 Search in Graphs	60
2.8 Summary	62
3. Unification-Based Formalisms for Translation in Natural Language Processing	65
3.1 Unification-Based Formalisms for Natural Language Processing	65
3.1.1 Definition of Typed Feature Structures with Appropriateness	68
3.1.2 Type Lattices	68
3.1.3 Feature Structures	69
3.1.4 Functions as Values of Features	73
3.2 Unification-Based Machine Translation	73
3.3 Architecture and Implementation of the Formalism	76
3.3.1 Definition and Implementation of Type Lattices	79
3.3.2 Definition and Implementation of Feature Structures	80
3.4 Summary	84

4. MILC: Structure and Implementation	85
4.1 Layered Charts	86
4.2 Communication Within the Application	95
4.2.1 Communication Architecture of an Application	96
4.2.2 Channel Models	98
4.2.3 Information Service and Synchronization	100
4.2.4 Termination	104
4.3 Overview of the Architecture of MILC	105
4.4 Word Recognition	106
4.5 Idiom Processing	108
4.6 Parsing	110
4.6.1 Derivation of Verbal Complexes	111
4.6.2 Spontaneous Speech and Word Recognition	113
4.6.3 Structure and Processing Strategies	115
4.7 Utterance Integration	121
4.8 Transfer	128
4.8.1 Chart-Based Transfer	130
4.8.2 The Implementation of Transfer for MILC	132
4.9 Generation	137
4.10 Visualization	143
4.11 Extensions	145
4.11.1 Extension of the Architecture	147
4.11.2 Anytime Translation	149
4.12 System Size	152
4.13 Summary	152
5. Experiments and Results	155
5.1 Hypergraphs	156
5.2 Translation	158
5.2.1 Data Material	158
5.2.2 Linguistic Knowledge Sources	159
5.2.3 Experiments and System Parameters	161
5.2.4 Evaluation	162
5.2.5 Extensions	164
5.3 Comparison With Non-incremental Methods	165
5.4 Summary	167
6. Conclusion and Outlook	169
Bibliography	175
Glossary	193
Index	195

List of Figures

1.1	The architecture of a functional unit within a modular system	4
1.2	The architectonic overview of MILC	17
1.3	Interlingua and transfer	19
1.4	Multi-level transfer	21
2.1	A chart for <i>Der Peter singt mit Freude</i>	29
2.2	The graph $K_{(3,3)}$	29
2.3	Word graph for the utterance n002k000	33
2.4	Unique word graph for the utterance n002k000	35
2.5	A difficult graph w.r.t the reduction to unique label sequences	40
2.6	Runtime for reducing graphs to unique label sequences	42
2.7	A complex graph for rank computation	46
2.8	Merging of silence edges	49
2.9	Two families of edges in a word graph	52
2.10	An interval graph	53
2.11	Two families of word hypotheses as hyperedges	54
2.12	Adding a word hypothesis to a hyperedge	58
2.13	Creation of additional paths by using hypergraphs	60
3.1	Feature structure for a simple syntactic rule	68
3.2	A transfer rule in LFG style	74
3.3	A transfer rule in TFS style	75
3.4	A transfer rule by Beskow	75
3.5	A small part of a type lattice	79
3.6	One part of the structure of lexical items	79
3.7	Feature structure for a simple syntactic rule	81
3.8	Definition of vertices within feature structures	82
4.1	The <i>Whiteboard-Architecture</i>	89
4.2	The principle layout of a layered chart	91
4.3	Principle component layout	97
4.4	The configuration of <i>split channels</i>	99
4.5	An example of a configuration file for <i>split channels</i>	101
4.6	Time line of the initial configuration of channels with the ILS	103
4.7	XPVM snapshot of the initial synchronization	104

4.8	The architectonic overview of MILC	106
4.9	A word graph in Verbmobil syntax	107
4.10	Idiom definition for the partial utterance "tut mir leid" ("I am sorry") ...	109
4.11	A syntactic rule for verbs in last position with four complements	112
4.12	One of the lexical entries for "Arbeitstreffen" ("work meeting")	116
4.13	A grammar rule for noun phrases with a determiner	118
4.14	A noun phrase from the partial parser	120
4.15	Island analysis: An example rule for nominal phrases	122
4.16	A rule for complements to the left of the verb	124
4.17	Lexicon entry for "ausmachen" ("schedule")	126
4.18	A verbal phrase from the integrator	127
4.19	A transfer chart for structural translations	132
4.20	A transfer rule for verbal phrases	133
4.21	The transfer lexicon entry for "recht sein" ("suit")	134
4.22	Topology of the fundamental rule for transfer	135
4.23	A subpart of the generation input for "lassen Sie uns das nächste Arbeitstreffen vereinbaren" ("let us schedule the next work meet- ing")	139
4.24	Generation rules for imperative verbs	140
4.25	Generation lexicon entry for "let"	142
4.26	A snapshot of the processing with MILC	145
5.1	Reduction of word edges by hypergraph conversion	156
5.2	Reduction of chart edges by hypergraph conversion	157
5.3	Reduction of analysis time by hypergraph conversion	158
5.4	Comparison of incremental and non-incremental processing	167

List of Tables

3.1	The syntax for type lattices	80
3.2	The syntax of feature structures	82
3.3	Internal representation of a feature structure	83
4.1	Messages sent from the word recognizer to the successor components . . .	108
4.2	Results for standard and partial parsing of the dialog n002k	115
4.3	Sketch of the processing of "lassen Sie uns den nächsten Termin ausmachen" ("let us schedule the next appointment")	125
4.4	An example for the output of the generator	143
4.5	System size (in lines of code)	153
5.1	Properties of the dialogs used in the experiments	159
5.2	The utterances in dialog m123	160
5.3	Results of analyzing five dialogs	162
5.4	Generator output for the utterance j534a005	163
5.5	Evaluation of the translations	164
5.6	Comparison of runtime for incremental and non-incremental configurations	165

List of Algorithms

1	Computing the topological order of a DAG	36
2	Computation of the transcript independent density of a word graph	37
3	Calculation of the number of paths in a graph	38
4	Reducing a graph to unique label sequences	39
5	Merging of two vertices	41
6	Calculation of the number of derivation steps of a fictitious parser for a word graph	43
7	Computing the rank of a path (part 1)	44
8	Computing the rank of a path (part 2)	46
9	Removing isolated silence edges	49
10	Removing consecutive silence	50
11	Removing all silence edges	51
12	Adding a word hypothesis e_n to a hypergraph $G = (\mathcal{V}, \mathcal{E}, \mathcal{L}, \mathcal{W})$	58
13	SSSP for DAGs	61
14	SSSP for incremental hypergraphs	62