

Text generation

**Using discourse strategies and focus constraints
to generate natural language text**

KATHLEEN R. McKEOWN

Department of Computer Science, Columbia University



Contents

Preface	ix
1. Introduction	1
1.1 Problems in generating text	1
1.2 A processing model	5
1.3 A sketch of related work	6
1.4 A text generation theory and method	8
1.5 System overview	11
1.6 The database application	13
1.7 Other issues	16
1.8 Guide to remaining chapters	17
2. Discourse Structure	19
2.1 Rhetorical predicates	20
2.1.1 Linguistic background	20
2.1.2 Ordering communicative techniques	21
2.2 Analysis of texts	24
2.2.1 Predicate recursiveness	30
2.2.2 Summary of text analysis	37
2.3 Related research using rhetorical predicates	38
2.4 Use of schemata	38
2.4.1 Associating technique with purpose	40
2.5 Selecting a schema	42
2.6 Filling the schema	45
2.7 An example	47
2.8 Future work	52
2.9 Conclusions	53
3. Focusing in discourse	55
3.1 Computational theories and uses of focusing	56
3.1.1 Global focus	56
3.1.2 Immediate focus	57
3.2 Focusing and generation	58
3.2.1 Global focus and generation	59
3.2.2 Immediate focus and generation	60
3.2.3 Current focus versus potential focus list	63
3.2.4 Current focus versus focus stack	65
3.2.5 Other choices	67

3.2.6 A focus algorithm for generation	69
3.2.7 Selecting a default focus	69
3.2.8 Overriding the default focus	70
3.2.9 The focus algorithm	71
3.2.10 Use of focus sets	73
3.3 Focus and syntactic structures	75
3.3.1 Linguistic background	75
3.3.2 Passing focus information to the tactical component	77
3.4 Future work	79
3.5 Conclusions	80
4. TEXT system Implementation	83
4.1 System components	84
4.2 Knowledge representation	87
4.2.1 Representation overview	87
4.2.2 Portability	90
4.2.3 Summary	91
4.2.4 The entity-relationship model	92
4.2.5 Use of generalization	93
4.2.6 The topic hierarchy	97
4.2.7 Relations	100
4.2.8 Distinguishing descriptive attributes	102
4.2.9 DDAs for database entity generalizations	103
4.2.10 Supporting database attributes	104
4.2.11 Based database attributes	106
4.2.12 DDAs for database entity subsets	109
4.2.13 Constant database attributes	111
4.3 Selection of relevant knowledge	113
4.3.1 Requests for information and definitions	113
4.3.2 Comparisons	114
4.3.3 Determining closeness	114
4.3.4 Relevancy on the basis of conceptual closeness	116
4.3.5 Conclusions	121
4.4 Schema implementation	122
4.4.1 Arc types	122
4.4.2 Arc actions	123
4.4.3 Registers	123
4.4.4 Graphs used	124
4.4.5 Traversing the graph	124
4.4.6 The compare and contrast schema	130
4.5 The tactical component	133
4.5.1 Overview of functional grammar	133
4.5.2 The grammar formalism	134
4.5.3 A functional grammar	138
4.5.4 The unifier	140
4.5.5 The TEXT system unifier	141

4.5.6 Unifying a sample input with a sample grammar	144
4.5.7 Grammar implementation	147
4.5.8 Morphology and linearization	151
4.5.9 Extensions	152
4.5.10 Disadvantages	152
4.5.11 Advantages	153
4.6 The dictionary	155
4.6.1 Design	155
4.6.2 Structure of dictionary entries	156
4.6.3 General entries	157
4.6.4 An example	160
4.6.5 Creating the dictionary	164
4.6.6 Conclusions	167
4.7 Practical considerations	168
4.7.1 User needs	168
4.7.2 Question coverage	169
4.7.3 Conclusions	169
5. Discourse history	171
5.1 Possible discourse 'history records	171
5.2 Questions about the difference between entities	172
5.3 Requests for definitions	178
5.4 Requests for information	180
5.5 Summary	183
6. Related generation research	185
6.1 Tactical components - early systems	186
6.2 Tactical components - later works	187
6.3 Generation in database systems	188
6.4 Planning and generation	189
6.5 Knowledge needed for generation	190
6.6 Text generation	190
7. Summary and conclusions	195
7.1 Discourse structure	195
7.2 Relevancy criterion	195
7.3 Discourse coherency	196
7.4 Generality of generation principles	196
7.5 An evaluation of the generated text	197
7.6 Limitations of the implemented system	199
7.7 Future directions	200
7.7.1 Discourse structure	200
7.7.2 Relevancy	203
7.7.3 Coherency	203
7.7.4 User model	204
7.7.5 Conclusion	204
Appendix A. Sample output of the TEXT system	205
Appendix B. Introduction to <i>Working</i>	223

Appendix C. Resources used	225
Appendix D. Predicate Semantics	227
Bibliography	237
Index	244