

Table of Contents

List of Figures	ix
Glossary of terms	xi
Introduction	1
1 The Framework : Scope and Concepts	3
1 Systems	3
1.1 Language processing	3
1.2 Systems and subsystems	6
1.3 System settings	11
1.4 NLP system example	15
2 Evaluation	19
2.1 Evaluation levels	19
2.2 Information retrieval experience	20
2.3 Applicability to NLP	28
2.4 NLP evaluation example	30
2.5 Speech processing illustrations	46
2.6 Evaluating generic systems	49
2.7 Evaluation tools	55
2.8 Evaluation from the social science point of view	59
2 NLP Evaluation : Work and Strategies	65
1 Evaluation so far	70
1.1 Machine translation (MT)	70
1.2 Message understanding (MU)	87
1.3 Database query (DBQ)	97
1.4 Speech understanding (SU)	109
1.5 Miscellaneous NLP task evaluations	115
1.6 Text retrieval	118
1.7 Cross-task issues	120
2 General developments	125
2.1 Evaluation workshops	125
2.2 Evaluation tutorials	155
2.3 EAGLES	157
2.4 Particular methodologies	163
2.5 Corpora, test suites, test collections, and toolkits	167
2.6 Generic NLP system evaluation	180
2.7 Mega-evaluation	185
2.8 Speech evaluation	188
3 Conclusions on evaluation to date	189

3 Strategies for Evaluation	193
1 General recommendations	193
1.1 Question framework	194
2 Evaluation illustration	196
2.1 Illustrative examples: scene setting	197
2.2 Examples: evaluations	201
3 Conclusion	218
References	219
Index	226