



# **An Introduction to Machine Translation**

**W. John Hutchins**

The Library,  
University of East Anglia,  
Norwich, UK

and

**Harold L. Somers**

Centre for Computational Linguistics,  
University of Manchester Institute of  
Science and Technology,  
Manchester, UK



**ACADEMIC PRESS**

*Harcourt Brace Jovanovich, Publishers*

LONDON SAN DIEGO NEW YORK BOSTON  
SYDNEY TOKYO TORONTO

This book is printed on acid-free paper

ACADEMIC PRESS LIMITED  
24-28 Oval Road  
LONDON NW1 7DX

*United States Edition published by*  
ACADEMIC PRESS INC.  
San Diego, CA 92101

Copyright © 1992 by  
ACADEMIC PRESS LIMITED

*All Rights Reserved*

No part of this book may be reproduced in any form by photostat,  
microfilm, or by any other means, without written permission  
from the publishers

A catalogue record for this book is  
available from the British Library

ISBN 0-12-362830-X

Printed in Great Britain at the University Press, Cambridge

# Contents

---

<b>Foreword</b> by Martin Kay .....	xi
<b>Preface</b> .....	xv
<b>List of abbreviations and symbols</b> .....	xix
<b>1. General introduction and brief history</b> .....	1
1.1. The aims of MT .....	2
1.2. Some preliminary definitions.....	3
1.3. Brief history of MT .....	5
1.4. Further reading .....	9
<b>2. Linguistic background</b> .....	11
2.1. The study of language .....	11
2.2. Grammar .....	12
2.3. Phonology and orthography .....	13
2.4. Morphology and the lexicon.....	15
2.5. Syntax .....	16
2.5.1. Syntactic features and functions.....	17
2.5.2. Deep and surface structure.....	18
2.5.3. Predicate-argument structure .....	19
2.6. Semantics.....	19
2.7. Text relations .....	20
2.8. Representations .....	21
2.8.1. Dependency .....	21
2.8.2. Phrase structure.....	23
2.8.3. Feature-based representations .....	25
2.8.4. Canonical and logical form .....	27

2.9. Formal grammar and linguistic theory .....	28
2.9.1. Context free grammars and rewrite rules .....	28
2.9.2. Transformational rules .....	30
2.9.3. Traces and gaps.....	34
2.9.4. X theory .....	35
2.9.5. Valency grammar .....	36
2.9.6. Case grammar .....	38
2.9.7. Unification grammar .....	39
2.10. Influential theories and formalisms .....	41
2.10.1. Lexical Functional Grammar.....	41
2.10.2. Categorical grammar .....	42
2.10.3. Government and Binding.....	43
2.10.4. Generalized Phrase Structure Grammar.....	43
2.10.5. Semantic compositionality .....	44
2.11. Further reading.....	45
<b>3. Computational aspects .....</b>	<b>47</b>
3.1. Data and programs .....	47
3.2. Separation of algorithms and data.....	48
3.3. Modularity .....	50
3.4. System design .....	51
3.5. Problems of input and output .....	51
3.5.1. Interactive systems .....	51
3.5.2. 'Foreign' languages.....	52
3.6. Lexical databases .....	54
3.7. Computational techniques in early systems .....	55
3.8. Parsing .....	56
3.8.1. Top-down and bottom-up .....	57
3.8.2. Backtracking.....	58
3.8.3. Feature notations .....	59
3.8.4. Trees .....	60
3.8.5. Charts .....	62
3.8.6. Production systems.....	65
3.9. Unification .....	66
3.10. Further reading.....	67
<b>4. Basic strategies .....</b>	<b>69</b>
4.1. Multilingual versus bilingual systems.....	69
4.2. Direct systems, transfer systems and interlinguas .....	71
4.3. Non-intervention vs. on-line interactive systems .....	77
4.4. Lexical data .....	78
4.5. Further reading .....	80
<b>5. Analysis .....</b>	<b>81</b>
5.1. Morphology problems.....	82
5.2. Lexical ambiguity .....	85
5.2.1. Category ambiguity.....	85
5.2.2. Homography and polysemy .....	86
5.2.3. Transfer ambiguity .....	87

5.3. Structural ambiguity .....	88
5.3.1. Types of structural ambiguity .....	88
Real structural ambiguity.....	88
Accidental structural ambiguity .....	89
5.3.2. Resolution of structural ambiguity .....	91
Use of linguistic knowledge .....	91
Contextual knowledge.....	92
Real world knowledge .....	93
Other strategies .....	94
5.4. Anaphora resolution.....	95
5.5. Quantifier scope ambiguity.....	96
5.6. Further reading.....	97
<b>6. Problems of transfer and interlingua.....</b>	<b>99</b>
6.1. Lexical differences .....	99
6.2. Structural differences .....	103
6.3. Levels of transfer representation .....	106
6.4. Morphological transfer .....	107
6.5. Transfer-based systems .....	109
6.5.1. Lexical transfer .....	113
6.5.2. Structural transfer.....	113
6.6. Transfer with a structural interlingua .....	116
6.7. Interlingua-based systems .....	118
6.7.1. Structural representation in interlingua systems.....	119
6.7.2. Lexical representation in interlingua systems .....	122
6.7.3. 'Restricted' interlinguas .....	124
6.8. Knowledge-based methods .....	124
6.9. Example-based methods .....	125
6.10. Summary: Comparison of transfer-based and interlingua-based systems.....	127
6.11. Further reading.....	129
<b>7. Generation.....</b>	<b>131</b>
7.1. Generation in direct systems.....	131
7.2. Generation in indirect systems.....	132
7.2.1. Generation in transfer-based systems .....	133
7.2.2. Generation in interlingua systems .....	136
7.3. Pre-determined choices and structure preservation .....	137
7.4. Stylistic improvements in generation .....	140
7.5. Generation as the 'reverse' of analysis .....	143
7.6. Further reading.....	145
<b>8. The practical use of MT systems .....</b>	<b>147</b>
8.1. Fully automatic high quality translation (FAHQT) .....	148
8.2. Machine-aided human translation (MAHT) .....	149
8.3. Human-aided machine translation (HAMT) .....	150
8.3.1. Pre-editing.....	151
8.3.2. Post-editing .....	152
8.3.3. Interactive MT .....	153

8.3.4. Interactive systems for monolingual users.....	154
8.4. Sublanguage systems .....	155
8.5. Use of low quality MT .....	157
8.6. Further reading .....	158
<b>9. Evaluation of MT systems .....</b>	<b>161</b>
9.1. Types and stages of evaluation .....	162
9.2. Linguistic evaluation of 'raw' output.....	163
9.2.1. Quality assessment .....	163
9.2.2. Error analysis .....	164
9.3. Evaluation by researchers.....	165
9.4. Evaluation by developers .....	166
9.5. Evaluation by potential users.....	166
9.5.1. Linguistic assessment .....	168
9.5.2. Evaluation of limitations, improvability, and extendibility.....	169
9.5.3. Technical assessment.....	170
9.5.4. Assessment of personnel requirements and ease of use ..	171
9.5.5. Evaluation of costs and benefits.....	171
9.6. Evaluation by translators .....	172
9.7. Evaluation by recipients .....	173
9.8. Further reading .....	174
<b>10. Systran.....</b>	<b>175</b>
10.1. Historical background .....	175
10.2. The basic system .....	177
10.2.1. Dictionaries .....	178
10.2.2. Computational aspects .....	179
10.2.3. Translation processes.....	180
10.3. Characteristics of the system .....	183
10.4. Improvability .....	186
10.5. Evaluations, users, and performance.....	187
10.6. Sources and further reading .....	189
<b>11. Susy.....</b>	<b>191</b>
11.1. Background .....	191
11.2. Basic system design.....	192
11.3. Data structure .....	194
11.4. Pre-editing and the fail-soft RESCUE operator .....	196
11.5. Analysis .....	196
11.5.1. Text input and dictionary look-up .....	196
11.5.2. Morphological analysis.....	197
11.5.3. Homograph disambiguation .....	198
11.5.4. Phrasal analysis .....	199
11.5.5. Structural analysis .....	200
11.5.6. Semantic disambiguation .....	201
11.6. Transfer and synthesis.....	202
11.7. Conclusion .....	204
11.8. Sources and further reading .....	205

<b>12. Météo</b> .....	207
12.1. Historical background .....	207
12.2. The translation environment: input, pre-processing and post-editing .....	208
12.3. The translation processes .....	209
12.3.1. Dictionary look-up .....	210
12.3.2. Syntactic analysis .....	212
12.3.3. Syntactic and morphological generation .....	215
12.4. The computational processes .....	215
12.4.1. The data structure .....	216
12.4.2. Rule formalism .....	217
12.4.3. Rule application .....	218
12.5. Summary and discussion .....	218
12.6. Sources and further reading .....	220
<b>13. Ariane (GETA)</b> .....	221
13.1. Historical background .....	221
13.2. General description .....	222
13.3. Multi-level representation .....	224
13.4. Linguistic processes .....	226
13.4.1. Morphological analysis .....	226
13.4.2. Multi-level analysis .....	226
13.4.3. Transfer and generation .....	229
13.5. Rule-writing formalisms .....	231
13.5.1. ATEF .....	232
13.5.2. ROBRA .....	233
13.5.3. TRANSF and SYGMOR .....	236
13.6. Concluding remarks .....	237
13.7. Sources and further reading .....	238
<b>14. Eurotra</b> .....	239
14.1. Background .....	239
14.2. Organisation and system design .....	241
14.3. Computational approach .....	244
14.3.1. Objects and structures .....	245
14.3.2. Translators and generators .....	245
14.3.3. Implementation .....	249
14.4. Linguistic aspects .....	249
14.4.1. Research on linguistic topics .....	249
14.4.2. An illustrative example .....	251
14.5. Conclusions .....	255
14.6. Sources and further reading .....	258
<b>15. METAL</b> .....	259
15.1. Historical background .....	259
15.2. The basic system .....	260
15.3. The linguistic databases .....	262
15.3.1. Dictionaries .....	262
15.3.2. Grammatical rules .....	266



15.4. The translation programs .....	271
15.5. Characteristics of the German-English system .....	275
15.6. Recent developments towards a multilingual system .....	276
15.7. Evaluations, users and performance .....	277
15.8. Sources and further reading .....	278
<b>16. Rosetta .....</b>	<b>279</b>
16.1. Background .....	279
16.2. Montague grammar .....	280
16.3. Reversibility and isomorphism .....	282
16.4. Translation processes .....	283
16.5. Structural correspondences .....	288
16.6. Subgrammars .....	290
16.7. Rule classes .....	291
16.8. Lexical transfer .....	292
16.9. Comments and conclusions .....	293
16.10. Sources and further reading .....	296
<b>17. DLT .....</b>	<b>297</b>
17.1. Background .....	297
17.2. The interlingua .....	298
17.3. System design .....	299
17.4. Dependency parsing .....	302
17.5. Metataxis .....	303
17.6. Interlingual data and SWESIL .....	305
17.6.1. English to Esperanto semantic processing .....	306
17.6.2. Esperanto to French semantic processing .....	308
17.6.3. Evaluation of SWESIL .....	310
17.7. Conclusions .....	311
17.8. Sources and further reading .....	311
<b>18. Some other systems and directions of research .....</b>	<b>313</b>
18.1. AI and Knowledge-based MT at CMU .....	313
18.2. Example-based MT at BSO .....	317
18.3. Statistics-based MT at IBM .....	320
18.4. Sublanguage translation: TITUS .....	322
18.5. MT for monolingual users .....	324
18.6. Speech translation: British Telecom and ATR .....	326
18.7. Reversible grammars .....	327
18.8. Computational developments .....	329
18.9. Concluding comments .....	330
18.10. Sources and further reading .....	332
<b>Bibliography .....</b>	<b>335</b>
<b>Index .....</b>	<b>351</b>

