

Machine Translation: Technologies and Applications

Volume 2

Series Editor:

Andy Way, *ADAPT Centre, Dublin City University, Ireland*

Editorial Board:

Sivaji Bandyopadhyay, *Jadavpur University, India*

Marcello Federico, *Fondazione Bruno Kessler, Italy*

Mikel Forcada, *Universitat d'Alacant, Spain*

Philipp Koehn, *Johns Hopkins University, USA and University of Edinburgh, UK*

Qun Liu, *Dublin City University, Ireland*

Hiroshi Nakaiwa, *Nagoya University, Japan*

Khalil Sima'an, *Universiteit van Amsterdam, The Netherlands*

François Yvon, *LIMSI, CNRS, Université Paris-Saclay, France*

This book series tackles prominent issues in MT at a depth which will allow these books to reflect the current state-of-the-art, while simultaneously standing the test of time. Each book topic will be introduced so it can be understood by a wide audience, yet will also include the cutting-edge research and development being conducted today.

Machine Translation (MT) is being deployed for a range of use-cases by millions of people on a daily basis. Google Translate and FaceBook provide billions of translations daily across many language. Almost 1 billion users see these translations each month. With MT being embedded in platforms like this which are available to everyone with an internet connection, one no longer has to explain what MT is on a general level. However, the number of people who really understand its inner workings is much smaller.

The series includes investigations of different MT paradigms (Syntax-based Statistical MT, Neural MT, Rule-Based MT), Quality issues (MT evaluation, Quality Estimation), modalities (Spoken Language MT) and MT in Practice (Post-Editing and Controlled Language in MT), topics which cut right across the spectrum of MT as it is used today in real translation workflows.

More information about this series at <http://www.springer.com/series/15798>

Bernard Scott

Translation, Brains and the Computer

A Neurolinguistic Solution to Ambiguity
and Complexity in Machine Translation



Springer

Bernard Scott
Tarpon Springs, FL, USA

ISSN 2522-8021 ISSN 2522-803X (electronic)
Machine Translation: Technologies and Applications
ISBN 978-3-319-76628-7 ISBN 978-3-319-76629-4 (eBook)
<https://doi.org/10.1007/978-3-319-76629-4>

Library of Congress Control Number: 2018935124

© Springer International Publishing AG, part of Springer Nature 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature.

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Gratefully Dedicated
A. M. D. G.

*Any intelligent fool can make things bigger
and more complex. It takes a touch of
genius – and a lot of courage – to move in
the opposite direction.*

– Albert Einstein

Preface

This book is about Machine Translation (MT) and the classic problems associated with this language technology. It is intended for anyone who wonders what if anything might be done to relieve these difficulties. For linguistic, rule-based systems, we attribute the cause of these difficulties to language's ambiguity and complexity and to their interplay in logic-driven processes. For non-linguistic, data-driven systems, we attribute translation shortcomings to the very lack of linguistics. We then propose a demonstrable way to relieve these drawbacks in both instances.

Throughout the book, we present a variety of translations by several of the most prominent linguistic, statistical and neural net MT systems in use today. Our object in doing this is to illustrate both the relative strengths and weaknesses of the various technologies these systems embody. The book's principal intent, however, is not to promote one particular translation system over against others (not even the one the author worked on for thirty years, described herein as Logos Model), but rather to examine the deeper and more critical question of the mechanisms that underlie the translation act itself, and to illustrate what can be done to optimize these mechanisms in a translation machine. We hold this to be the more fundamental issue that needs to be addressed if the classic problems associated with MT are to be solved, and consistent, high-quality machine output is ever to be realized.

Because the linguistic processes of the brain are singularly free of the classic difficulties that beset the machine, we have looked to the brain for possible guidance. We describe a working translation model (Logos Model) that has taken its inspiration from key assumptions about psycholinguistic and neurolinguistic function. We suggest that this brain-based mechanism is effective precisely because it bridges both linguistically-driven and data-driven methodologies. In particular, we show how simulation of this cerebral mechanism has freed this one MT effort, Logos Model, from the all-important, classic problem of complexity when coping with the ambiguities of language. Logos Model accomplishes this by a data-driven process that does not sacrifice linguistic knowledge, but that, like the brain, integrates linguistics within a data-driven process. As a consequence, we suggest that the brain-like mechanism simulated in this model has the potential to contribute to further advances in MT in all its technological instantiations.

These admittedly are controversial claims, and we recognize that the reader may be inclined to dismiss them out of hand, especially given the fact that the model being described, Logos Model, had its origins more than 45 years ago in the earliest days of MT. How, one will ask, can technology from so far back in time offer anything of interest to present-day MT? That is certainly a legitimate question, but it is one we trust this book will answer.

As readers work their way through this book, they will see that we are showing Logos Model at its best, seemingly at times at the expense of other translation systems. Our purpose in writing this book, however, has *not* been to prove that Logos is a better translation system. In terms of the general output quality of many MT systems nowadays, no such claim could be defended. Our purpose rather has been quite different, namely, to demonstrate that the technology underlying Logos Model offers a demonstrable solution to the problem that complexity poses for MT. As we argue throughout this book, complexity is the one issue that is most apt to limit the ultimate potential of any MT system, whether linguistic, statistical or neural. And we attempt to show that Logos Model, originally designed as it was to address the complexity problem, may offer a workable answer. Logos Model translations shown in this book are meant to demonstrate that the model must be doing something right in that regard, something that we trust would be of interest to MT developers generally. Allow me to repeat the point. Logos Model translations in this book are *not* intended to prove that Logos is a better system, only that underlying Logos Model technology may have something of genuine interest to offer. It is hoped that the MT community will understand this, and that the empirical data, arguments and personal testimony we have presented will be considered in constructive spirit intended.

One final matter. Translations shown in this book by Google Translate, Microsoft's Bing Translator, SYSTRANet, PROMT Translator and LISA Lab's neural MT system were carried out in the 2016–2017 timeframe. Readers should be aware that these translations do not necessarily represent output of these systems subsequent to this 2016–2017 timeframe. Readers will note that output from Google Translate and Bing Translator had to be marked as either statistical or neural, since both the Google and Microsoft systems transitioned to neural net technology in late 2016 as this book was being written.

Tarpon Springs, FL, USA

Bernard Scott

Acknowledgements

The following individuals, among still others, are responsible for having turned the concepts of this book into a working system.

Linguists

Kutz Arrieta
Anabela Barreiro
Mary Lou (Christine) Faggione
Claudia Gdaniec
Brigitte Orliac
Karen (Liz) Purpura
Patty Schmid
Steve Schneiderman
Sheila Sullivan
Monica Thieme

Systems

Charles Byrne (Logos co-founder)
Daniel Byrne
Jonathan Lewis
Oliver Mellet
Barbara Tarquinio
Kevin Wiseman
Theresa Young

Corporate

Cecilia Mellet
Robert Markovits
Arnold Mende
James Linen
Charles Staples
Hans Georg Forstner

Special thanks to Anabela Barreiro and Fernando Batista (INESC-ID) without whose assistance this book would not have been possible. And much appreciation for helpful suggestions by two anonymous reviewers.

Permissions

All figures in this book were created by the author except for those listed below, used with permissions indicated:

Fig. 4.3 Permission granted by Dr. Francis Wong

Fig. 4.5 Permission granted by National Institute of Health under GNU Free Document License (GFDL)

Fig. 4.6 Permission granted by Oxford University Press

Fig. 5.1 Permission granted by Elsevier

Postscript Fig. 5G-1 Permission granted by World Scientific Publishing

Fig. 8.1 Permission granted by the Association of Computational Linguistics (ACL)

Fig. 8.2 Permission granted by Christoph Bugmer under GNU Free Document License (GFDL)

Fig. 8.4 Permission granted by Elsevier

Contents

Part I

1	Introduction	3
	References.....	11
2	Background	13
2.1	Logos Model Beginnings	13
2.2	The Advent of Statistical MT.....	16
2.2.1	Pattern-Based Processes in SMT and Logos Model	17
2.3	Overview of Logos Model Translation Process	20
2.4	Psycholinguistic and Neurolinguistic Assumptions	25
2.4.1	First Assumption	25
2.4.2	Second Assumption.....	26
2.4.3	Third Assumption.....	26
2.5	On Language and Grammar	27
2.5.1	The Origin and Nature of Grammar.....	28
2.5.2	Language, Grammar and Associative Memory.....	29
2.5.3	In Principio Erat Verbum.....	30
2.6	A Note About Neural MT (NMT).....	32
2.7	Conclusion.....	33
	Postscripts	34
	Postscript 2-A.....	34
	Postscript 2-B	37
	Postscript 2-C	37
	References.....	38
3	Language and Ambiguity: Psycholinguistic Perspectives	41
3.1	Levels of Ambiguity.....	41
3.2	Language Acquisition and Translation.....	44
3.2.1	Linguistic Processes Involved in Second Language Acquisition	45
3.2.2	On Learning French	49

3.3	Psycholinguistic Bases of Language Skills.....	51
3.3.1	Analytical Basis	51
3.3.2	Empirical Basis	52
3.3.3	Analogical Basis.....	52
3.4	Practical Implications for MT	54
3.4.1	Semantico-Syntactic Solutions to the Problem of Ambiguity in MT	55
3.5	Psycholinguistics in a Machine.....	58
3.5.1	Generate Target Translation	61
3.5.2	OpenLogos and Logos Model.....	61
3.6	Conclusion.....	63
	References.....	63
4	Language and Complexity: Neurolinguistic Perspectives	65
4.1	On Cognitive Complexity	65
4.2	A Role for Semantic Abstraction and Generalization	70
4.3	Connectionism and Brain Simulation	73
4.4	Logos Model As a Neural Network	76
4.5	Language Processing in the Brain.....	79
4.5.1	Cortical Circuits and Logos Model.....	80
4.5.2	Hippocampus and Logos Model	81
4.6	MT Performance and Underlying Competence	89
4.7	Conclusion.....	91
	Postscripts	91
	Postscript 4-A.....	91
	Postscript 4-B	92
	Postscript 4-C	94
	Postscript 4-D.....	94
	Postscript 4-E	94
	Postscript 4-F	95
	Postscript 4-G.....	95
	Postscript 4-H.....	96
	References.....	96
5	Syntax and Semantics: Dichotomy Versus Integration	99
5.1	Syntax Versus Semantics: Is There a Third, Semantico-Syntactic Perspective?.....	99
5.2	Recent Views of the Cerebral Process.....	106
5.3	Syntax and Semantics: How Do They Relate?.....	108
5.4	Conclusion.....	113
	Postscripts	114
	Postscript 5-A.....	114
	Postscript 5-B	115
	Postscript 5-C	115
	Postscript 5-D.....	116

Postscript 5-E	117
Postscript 5-F	118
Postscript 5-G	119
References	123
6 Logos Model: Design and Performance	127
6.1 The Translation Problem	127
6.1.1 Five Fundamental Design Decisions	129
6.2 How Do You Represent Natural Language?	130
6.2.1 Effectiveness of SAL for Deterministic Parsing	131
6.3 How Do You Store Linguistic Knowledge?	133
6.3.1 General Remarks	133
6.3.2 Logos Model Lexicon	133
6.3.3 The Pattern-Rule Database	134
6.4 How Do You Apply Stored Knowledge to the Input Stream?	137
6.4.1 Modules RES1 and RES2 (R1 and R2)	138
6.4.2 Module PARSE1	140
6.4.3 Module PARSE2	142
6.4.4 Module PARSE3	145
6.4.5 Module PARSE4	149
6.5 How Do You Effect Target Generation?	153
6.5.1 Target Components	153
6.6 How Do You Cope with Complexity?	153
6.6.1 A Final Illustration	154
6.7 Conclusion	157
Postscripts	157
Postscript 6-A	157
Postscript 6-B	158
References	162
7 Some Limits on Translation Quality	163
7.1 First Example	164
7.2 Second Example	166
7.3 Other Translation Examples	167
7.4 Balancing the Picture	168
7.5 Conclusion	169
References	171
8 Deep Learning MT and Logos Model	173
8.1 Points of Similarity and Differences	174
8.2 Deep Learning, Logos Model and the Brain	179
8.3 On Learning	181
8.4 The Hippocampus and Continual Learning	186
8.5 Conclusion	193
8.6 A Final Demonstration	196
References	200

Part II

9	The SAL Representation Language	205
9.1	Overview of SAL	205
9.2	SAL Parts of Speech	206
9.2.1	Open Classes (Table 9.1).....	206
9.2.2	Closed Classes (Table 9.2)	206
9.3	SAL Nouns (WC 1).....	207
9.3.1	Aspective Nouns	209
9.3.2	Concrete Nouns	211
9.3.3	Animate Nouns.....	212
9.3.4	Abstract Nouns.....	213
9.3.5	Measure Nouns.....	214
9.3.6	Place Nouns.....	215
9.3.7	Mass Nouns.....	216
9.3.8	Information and Time Nouns	217
9.4	SAL Verbs (WC 2)	218
9.4.1	The Intransitive-Transitive Verb Spectrum.....	219
9.4.2	Intransitive Verbs.....	222
9.4.3	Subjective Transitive Verbs	223
9.4.4	Reciprocal Transitive Verbs.....	224
9.4.5	Ditransitive Verbs	225
9.4.6	Objective Transitive Verbs.....	226
9.4.7	Pre-process Verbs	227
9.4.8	Simple Preverbal Verbs	228
9.4.9	Preverbal Complex Verbs	229
9.4.10	Preverbal-Preclausal Verbs.....	230
9.4.11	Preclausal Verbs	231
9.5	SAL Adjectives (WC 4)	233
9.5.1	Preclausal/Preverbal Adjectives	234
9.5.2	Preverbal Adjectives.....	235
9.5.3	Adverbial Adjectives	236
9.5.4	Non-adverbial Adjectives	237
9.6	SAL Adverbs (WC 3 and WC 6).....	238
9.6.1	Locative Adverbs have the Following Supersets.....	238
9.6.2	Non-locative Adverb have the Following Supersets	239
	Postscript.....	240
	Postscript 9-A.....	240