



Automated Translation Technology

*Using Machine Translation to
Close the Translation Gap and
Fuel Information Discovery*

By Donald A. DePalma and
Robert J. Kuhns

November 2006

Automated Translation Technology

By Donald A. DePalma and Robert J. Kuhns

November 2006

ISBN: 1-933555-31-9

978-1-933555-31-9

Copyright © 2006 by Common Sense Advisory, Inc., Lowell, Massachusetts,
United States of America.

Published by:

Common Sense Advisory, Inc.
100 Merrimack Street
Suite 301
Lowell, MA 01852-1708 USA
+1.866.510.6101 or +1.978.275.0500
info@commonsenseadvisory.com
www.commonsenseadvisory.com

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher.

Permission requests should be addressed to the Permissions Department, Common Sense Advisory, Inc., Suite 301, 100 Merrimack Street, Lowell, MA 01852-1708, +1.978.275.0500, E-Mail: info@commonsenseadvisory.com. See www.commonsenseadvisory.com/en/citationpolicy.html for usage guidelines.

Trademarks: Common Sense Advisory, Global DataSet, DataPoint, Globa Vista, Quick Take, and Technical Take are trademarks of Common Sense Advisory, Inc. All other trademarks are the property of their respective owners.

Information is based on the best available resources at the time of analysis. Opinions reflect the best judgment of Common Sense Advisory's analysts at the time, and are subject to change.

Table of Contents

Topic.....	1
Using Machine Translation to Close the Translation Gap	1
Structure and Navigation of the Report	1
Who Should Read This Report?.....	2
Demand	3
The Choice: Human, Machine, or Zero Translation.....	3
Today More Web Users “Pull” MT than Receive Published MT Content	4
How Good Does Automated Translation Output Have to Be?	5
Market Demand for Local Language Causes Human Translation Gap	5
How and Where Organizations Deploy Automated Translation	7
Where You Can Find MT in 2006.....	8
More Ambitious Uses of MT Assume Friendly Audiences.....	13
What Does the Future Hold for Automated Translation Applications?	18
MT Market: Fifty Years in the Making, But Still Short of Expectations	19
Mismatch: Market Expectations versus Economic Value	20
The Business Side of Automated Translation	20
Strengths of Machine Translation	21
Weaknesses of Machine Translation	22
Opportunities for Machine Translation	23
Threats to Machine Translation.....	25
Conclusion	25
Buying Guide	27
The Techno-Religious Question: Rules, Statistics, or None of the Above?	27
Is Your Content Ready for Automated Translation?	28
Software License Costs Range from Pennies to Real Money	28
In Which Language Do You Need Your Content?	30
Where Do You Need Machine Translation to Run?.....	31
Integration with the Corporate Technology Stack.....	31
Integration with Global Content Life Cycle	33
How Can You Improve MT Output?	34
The Final Issue: Performance, Sure, But Your Mileage May Vary	36
Where Will Automated Translation Suppliers Go in the Next Few Years?.....	36
Conclusions from Evaluating These MT Vendors.....	38
Technology Probe.....	41
Automated Translation Covers Multiple Technologies	41
Rules-Based Systems: Pros and Cons	41
Statistical MT Systems: Pros and Cons.....	43
Hybrid MT Systems: Pros and Cons.....	44
Context-Based Systems: Pros and Cons	47
How MT Specialists Evaluate Automated Translation Software	48
BLEU (Bilingual Evaluation Understudy) from IBM Research	49
NIST Metric from the National Institute of Standards and Technologies.....	50
F-Measure from New York University	50
Which Measure Is Best?.....	50

Appendix.....	52
🌐 About Common Sense Advisory	55
🌐 Future Research.....	55

Figures

Figure 1: Half of Consumers Use MT to Understand Anglophone Websites.....	4
Figure 2: Zero Translation Dominates, With Most Content Never Getting Translated	6
Figure 3: Criteria for When HT and MT Are Appropriate	7
Figure 4: Before and After Free Online Translation – Pull MT	9
Figure 5: Some Organizations Minimally Alert Visitors to MT Use	10
Figure 6: Some Sites Highlight Potential Problems with MT.....	10
Figure 7: Some Companies Don't Tell Users that It's MT behind the Scenes	11
Figure 8: Some Companies Use MT for General Website Publishing.....	12
Figure 9: Much Information Inside Organizations Has More Value Once Translated	14
Figure 10: MT Can Populate Knowledge Bases for Well-Defined Domains.....	15
Figure 11: MT Can Help Educators When Bilingual Education Funding Falls Short .	15
Figure 12: Multilingual Search	16
Figure 13: Post-Edited MT Works Better for Customer Service	17
Figure 14: Chat Ensures Everyone Shows Up in Time for Coffee and Halwachmar ..	19
Figure 15: Revenue and Visibility Remain Below the Waterline for MT Suppliers	21

Tables

Table 1: Drivers for More Translation	6
Table 2: Is Your Content Ready for MT? Be Prepared for Tradeoffs	29
Table 3: Sample Rows from Language Support Tables in Appendix	30
Table 4: MT Support for Integration with Corporate Applications	32
Table 5: Web-Centric Markup Languages Dominate File Type Support	33
Table 6: Most MT Systems Have Weak Support for Content Life Cycle.....	34
Table 7: Dictionary and System Customizations	35
Table 8: Performance Varies Widely for Automated Translation Engines	37
Table 9: Development Plans for Automated Translation Solutions.....	39
Table 10: Summary Checklist for Evaluating Automated Translation Software	40
Table 11: Today's Commercial Off-the-Shelf Automated Translation Technologies ..	42
Table 12: Evolving Automated Translation Technologies.....	46
Table 13: Bi-directional European Language Pairs with English	52
Table 14: Bi-directional Non-European Language Pairs with English	53
Table 15: Bi-directional Language Pairs Not Involving English.....	53
Table 16: Single-Direction Language Pairs	54



Topic

Using Machine Translation to Close the Translation Gap

Content volume is increasing faster than any company or government can manage, much less translate into all the languages of their employees or external users. Faced with mass quantities of information written in other languages, most organizations choose not to translate most of that content (we call this “zero translation”). Many take the path of high-quality but relatively expensive human translation (HT), picking and choosing just a small subset of their corporate content to translate. A growing number choose the faster, cheaper, but imperfect option of machine translation (MT).

Machine translation is a computer application that analyzes text in a source language and produces an equivalent text in another tongue. Functionally speaking, MT parses text into parts of speech such as nouns, verbs, and adjectives. Then it processes these linguistic components according to linguistic rules, statistical algorithms, or a combination of these methods.

Of all natural language processing (NLP) technologies, automated translation harbors the most potential as a disruptive technology for global communications and commerce. In this report we advise companies operating internationally to evaluate MT as a way to accelerate content availability in other languages for both communication and commerce and to ensure that no content is left behind.

Structure and Navigation of the Report

This report consists of three parts, each aimed at helping information publishers determine the suitability of machine translation for their applications:

- **Demand.** This section will educate organizations considering the use of automated translation technology by answering questions such as “what is MT?” and “how can I determine whether it meets my needs?” It defines machine translation, describes the demands driving the development of automated translation, outlines how it is currently being used in industry and government, and categorizes the major MT features. We also estimate the size of the market for automated translation technology.
- **Buying Guide.** This part will be useful to organizations evaluating MT as a way to provide information to their customers or constituencies – either by

translating foreign-language content on-demand or by incorporating MT technology into their content life cycles. It outlines a step-wise approach for choosing an MT solution based on system architecture, language pair needs, standards, application programming interfaces for integration, and performance factors such as translation throughput.

- **Technology Probe.** This section is only for the stout of heart. Anyone truly interested in learning more about the science of automated translation technology will find value in our dissection of the four MT approaches most favored by developers. It describes each model, lists its advantages and disadvantages, and outlines the major metrics used by industry, academe, and government to assess the performance of these four approaches.

For details about solutions discussed in this report, see “Automated Translation Suppliers” (Nov06), a summary of the vendor research we conducted for this report. That report summarizes the automated translation products of nine independent software vendors (ISV) in the United States and Europe that represent a cross-section of the MT development community, including existing commercial off-the-shelf (COTS) software, research projects, a service offering, and one product that is quite new to the market.

Who Should Read This Report?

When used in conjunction with our “Automated Translation Suppliers” report, this document should be most useful to anyone evaluating the use of automated translation solutions. This report will not be useful to information consumers wondering whether they should use automated translation technology.

Typical evaluators include:

- **Publishers and other information disseminators.** Any organization that produces information for global consumption may find MT useful for filling in the gaps of human translation (HT) or increasing the effectiveness of their HT budgets. Evaluators, decision makers, information technologists, and content management specialists need this information to determine the suitability of MT for their applications and the ability to integrate automated translation into their technology infrastructure.
- **Language service providers.** Some LSPs use MT for pre-processing jobs. Evaluators will be interested in optimization and interfaces for integrating with their core operating and workflow systems.