



Deliverable D4.3

Survey on the State of HQMT in Industry and LSPs

Authors: Georg Rehm (DFKI), Aljoscha Burchardt (DFKI)

Dissemination Level: Public

Date: 08 August 2016

Status: Final



Grant agreement no.	645357
Project acronym	CRACKER
Project full title	Cracking the Language Barrier
Type of action	Coordination and Support Action
Coordinator	Dr. Georg Rehm (DFKI)
Start date, duration	1 January 2015, 36 months
Dissemination level	Public
Contractual date of delivery	31/03/2016
Actual date of delivery	08/08/2016 (overall approach and delay of this specific deliverable discussed with the project officer in March 2016)
Deliverable number	D4.3
Deliverable title	Survey on the State of HQMT in Industry and LSPs
Type	Report
Status and version	Final
Number of pages	37
Contributing partners	DFKI
WP leader	CUNI
Task leader	DFKI
Authors	Georg Rehm (DFKI), Aljoscha Burchardt (DFKI)
Internal reviewers	n.a. (overall approach regarding the survey and the two suggested phases discussed and agreed upon in a CRACKER consortium meeting in March 2016)
EC project officer	Pierre-Paul Sondag (M01-M18), Susan Fraser (M19-M36)
The partners in CRACKER are:	<ul style="list-style-type: none"> • Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), Germany • Charles University in Prague (CUNI), Czech Republic • Evaluations and Language Resources Distribution Agency (ELDA), France • Fondazione Bruno Kessler (FBK), Italy • Athena Research and Innovation Center in Information, Communication and Knowledge Technologies (ATHENA RC), Greece • University of Edinburgh (UEDIN), UK • University of Sheffield (USFD), UK

For copies of reports, updates on project activities, and other CRACKER-related information, contact:

DFKI GmbH
CRACKER

Dr. Georg Rehm
Alt-Moabit 91c
D-10559 Berlin, Germany

georg.rehm@dfki.de
Phone: +49 (0)30 23895-1833
Fax: +49 (0)30 23895-1810

Copies of reports and other material can also be accessed via <http://cracker-project.eu>.

© 2016 CRACKER Consortium

Contents

<u>1</u>	<u>Introduction</u>	<u>4</u>
<u>2</u>	<u>Original Approach to the Survey</u>	<u>4</u>
<u>3</u>	<u>Initial Preparation Work</u>	<u>5</u>
<u>4</u>	<u>The CommonSenseAdvisory Surveys (early 2016)</u>	<u>5</u>
<u>5</u>	<u>Collaboration between CommonSenseAdvisory and CRACKER</u>	<u>6</u>
5.1	Phase 1: Initial Pilot Collaboration (2016)	6
5.2	Phase 2: Comprehensive Survey (2017)	8
<u>6</u>	<u>The CSA Summary Report “Europe’s Leading Role in MT”</u>	<u>8</u>

1 Introduction

The Coordination and Support Action CRACKER is carrying out multiple different activities in the area of European Language Technologies (including Multilingual Technologies) such as, for example, community building, organising events (conferences, workshops, shared tasks), annotating data sets, providing tool support and preparing a Strategic Agenda for the Multilingual Digital Single Market.

CRACKER also contains a survey component to learn more about the current state of play and also uptake of technologies for high-quality machine translation (HQMT) in Europe, especially with regard to their economic impact and to identify informative success stories originating in European projects, European companies or European research. Through this survey, we intended to find out to what extent EC/EU-funded projects managed to provide and transfer to market any technologies, services or approaches (e.g., workflows etc.), that are now being used in production environments in industry and/or in LSPs (language service providers).

2 Original Approach to the Survey

The survey is specified in Task 4.1 of the CRACKER Description of Action:

In this task, we will perform a **survey of social and economic impact of MT tools, standards, and resources**, especially of recent **results of EU-funded research** (who uses which technologies and approaches, how, and with which impact?). The survey will target **LSPs** and (very) large **industry users** of translation. On the one hand, the survey will document the actual or planned **uptake of project results** including Moses, MateCat, Panacea, META-SHARE, ITS 2.0, MQM, etc. and, on the other, **promote these results** among the participants of the survey. The survey is planned for M15. If needed, a follow-up survey will be performed.

The technology area of Machine Translation is still suffering from a distorted image and lack of awareness on the side of decision-makers in business, politics and public administration. Earlier studies tried to estimate the markets and identify informative success stories but the dynamics of recent developments have made these surveys outdated. A concise account of the true existing impact of translation technologies on society and economy is missing. In the CRACKER proposal we suggested to outsource the preparation and operationalisation of the survey to one or more subcontractors, who would conduct this survey of social and economic impact of MT, especially of recent results of EU-funded research (users, workflows, impact).

An additional goal associated with the survey was that it would generate interest in MT and HQMT and help to create a more realistic image of the potential of MT. The survey was also supposed to support CRACKER's outreach and transfer strategy of (HQ)MT results, especially communication with existing and prospective users of MT through a close cooperation between the MT research community and LSPs.

According to the Description of Action, Task 4.1 was foreseen to start in M4 (April 2015) and to end in M15 (March 2016).

The final survey, CRACKER Deliverable D4.3, was described and further specified in the Description of Action as follows:

D4.3: Survey on the state of HQMT in industry and LSPs

This report summarises the results of the survey on the economic impact and uptake of recent EC-funded MT actions, especially with regard to industry and language service providers (LSPs).

For the preparation of the survey we had foreseen a subcontracting budget. Its target audience are business, politics and public administration.

3 Initial Preparation Work

While the collection of questions and ideas for the survey was started even before the actual planned start of the task, i.e., April 2015 (M4), the systematic preparation of the survey proper started in November 2015. The initial preparation work included the collection and structuring of questions as well as interesting themes and topics, the collection and structuring of contacts, companies, experts, and other stakeholders to be approached and also preparing a shortlist of potential subcontractors.

4 The CommonSenseAdvisory Surveys (early 2016)

The first potential subcontractor we contacted to discuss the survey was the US-based think tank and consulting company CommonSenseAdvisory (CSA), who are specialised on the areas of Internationalisation and Globalisation.

CRACKER had prepared a short presentation with the overall goals of the CRACKER survey, which was discussed in a first phone meeting with CSA in January 2016. In these initial discussions we learned that CSA was conducting two rather thorough and comprehensive online surveys: one on MT suppliers (containing a total of 59 questions), the other one on MT demand (containing 47 questions). Both surveys were already open when CRACKER approached CSA; the two surveys were scheduled to close on 17 February 2016.

After learning about the two CSA surveys, the team that was preparing the survey in CRACKER was immediately concerned about the severe danger that the CSA surveys would jeopardise the operationalisation and results of an independent CRACKER survey due to the significant overlap between the foreseen respondents and stakeholder groups addressed by the two CSA surveys and a separate CRACKER survey. The threat was that, once stakeholders had responded to the CSA surveys (which were already running), there was only a small chance that they would also respond, only a few weeks later, to a very similar second survey, the one conducted by CRACKER. In other words, the impact of the two CSA surveys on the CRACKER survey would have been that the CRACKER survey would – most probably – had received much less responses because target stakeholders are typically unlikely to respond to multiple surveys in a short period of time.

This is why CRACKER immediately suggested to discuss with CSA if we could find potential synergies between their ongoing work and the planned CRACKER survey.

5 Collaboration between CommonSenseAdvisory and CRACKER

Between January and March 2016, CRACKER and CommonSenseAdvisory had multiple phone meetings, in which we discussed and agreed upon the overall two-phase approach briefly sketched in the following two Sections.

Phase 1, the initial pilot collaboration (2016), relates to the present deliverable and report (see Section 5.1). Phase 2, a more comprehensive study (2017), relates to future work, planned for the third and last year of CRACKER (see Section 5.2).

To summarise the two-phase approach, instead of preparing and running our own survey as CRACKER in early 2016, CSA prepared – for CRACKER – a customised summary of their own MT survey report, which is tailored to the European landscape and CRACKER situation. CSA provided this summary report to CRACKER free of charge. The delivery of this summary report concluded Phase 1 of the collaboration in early July 2016.

Phase 2 is planned for 2017. One year after the previous CSA MT surveys, CRACKER and CSA have agreed to collaborate on a new, dedicated survey, which will be financially supported through the subcontract foreseen in CRACKER.

For Phase 1, the subcontracting budget foreseen in the CRACKER budget has not been made use of.

5.1 Phase 1: Initial Pilot Collaboration (2016)

Initially, in January 2016, CommonSenseAdvisory provided the full questionnaires of the two surveys that were then analysed by CRACKER. We found a significant overlap between the two CSA surveys and the questions CRACKER wanted to learn more about in our own survey (ca. 75%).

The CRACKER team explained, in detail, the current European situation, the original goal of the CRACKER survey, our main themes of interest and we also pointed out those questions from the two CSA surveys that are most relevant for CRACKER.

Furthermore, CRACKER and also other projects of the Cracking the Language Barrier federation (such as, for example, QT21) helped with the distribution of the two CSA survey links in order to maximise their coverage.

Based on the exhaustive briefing through the CRACKER team, CSA prepared a customised summary of their own MT survey report, which is tailored to the European landscape and situation.

In terms of the date of delivery, CRACKER and CSA agreed that this customised summary should be provided in early July 2016 so that it can be presented at META-FORUM 2016, which took place on 4/5 July 2016, in Lisbon, Portugal, and made available through the CRACKER website. CSA provided the summary on 1 July 2016. The CSA report was made available on the CRACKER website immediately after delivery (see Figure 1), see <http://cracker-project.eu/csa-mt-report/>.

It is important to stress that CSA prepared and provided this summary report to CRACKER fully free of charge, i.e., the subcontracting budget foreseen for the survey activity has not been made use of yet.

Regarding the availability of the summary report, the authors are happy to report that CRACKER has been able to negotiate with CSA that they apply a Creative Commons-based license for the summary report (Creative Commons Attribution-NoDerivatives 4.0 International, i.e., CC BY-ND 4.0). This license enables CRACKER to share the report publicly on its website (“Under this license, you are free to: Share — copy and redistribute the material in any medium or format for any purpose, even commercially.”)



Figure 1: The CSA report on the CRACKER website (<http://cracker-project.eu/csa-mt-report/>)

The delivery of this summary report concluded Phase 1 of the collaboration in early July 2016.

The full CSA report – Common Sense Advisory: “Europe’s Leading Role in Machine Translation. How Europe Is Driving the Shift to MT”, by Arle R. Lommel and Donald A. DePalma – is attached to this deliverable.

5.2 Phase 2: Comprehensive Survey (2017)

Phase 2 of the collaboration between CSA and CRACKER is foreseen for the first half of 2017. Approximately one year after the two previous CSA MT surveys, CRACKER and CSA will collaborate on a new, much more thorough and in-depth survey, which will be financially supported through the subcontract foreseen in CRACKER and which will, in terms of concept, scope, topics and questions, be fully customised towards the needs and demands of the European situation, also taking into account the recommendations expressed in CRACKER’s Strategic Research and Innovation Agenda and the political situation (especially the situation with regard to funding for research, development and innovation). The goal is to have the final report ready by the time META-FORUM 2017 takes place (date not yet specified).

6 The CSA Summary Report “Europe’s Leading Role in MT”

The CSA summary report “Europe’s Leading Role in Machine Translation”, by Arle R. Lommel and Donald A. DePalma is included below.



Europe's Leading Role in Machine Translation

*How Europe Is Driving
the Shift to MT*

By Arle R. Lommel and
Donald A. DePalma

June 2016

Produced for the CRACKER project



Europe's Leading Role in Machine Translation

By Arle R. Lommel and Donald A. DePalma

June 2016

Copyright © 2016 by Common Sense Advisory, Inc., Cambridge, Massachusetts, United States of America.

This report is made available under the Creative Commons Attribution-NoDerivatives 4.0 International (CC BY-ND 4.0) license.

Under this license, you are free to:

Share — copy and redistribute the material in any medium or format for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

- **Attribution** — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- **No Derivatives** — If you remix, transform, or build upon the material, you may not distribute the modified material.
- **No additional restrictions** — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Details of this license are available at:

<https://creativecommons.org/licenses/by-nd/4.0/legalcode>

Requests for permission to use this report or portions thereof outside the context permitted under the license above should be addressed to the Permissions Department, Common Sense Advisory, Inc., 100 Cambridgepark Drive, Cambridge, MA 02140, +1.978.275.0500, E-Mail: info@commonsenseadvisory.com. See www.commonsenseadvisory.com/en/citationpolicy.html for usage guidelines.

Trademarks: Common Sense Advisory, Global Watchtower, Global DataSet, DataPoint, Globa Vista, LSP Metrix, MarketFlex, Quick Take, Technical Take, and TMS Live are trademarks of Common Sense Advisory, Inc. All other trademarks, product names, logos, and brands 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

Executive Summary	1
Overview of Enterprise MT	2
The Role of MT in Europe	3
North America Leads in MT Demand	3
Europe Leads MT Production	4
Drivers for MT Adoption	5
Large Enterprises and LSPs Led the Way	5
Smaller Ones Are Catching Up	6
MT Gains Traction for High Volumes in European Languages	7
Increasing Translation Volumes Make MT Inevitable	8
MT Adoption Models	10
Enterprises Have Three Models for MT Adoption	10
Three Enterprise Models	10
MT Adoption Follows a Curve	11
The Curve Is Changing	12
LSPs Move to Adopt MT	12
Audiences and Content Types	15
Enterprises Aim MT Output at Underserved Customers	15
Business Content Leads Machine Translation	16
Customer Engagement Content Remains Largely Untranslated	17
MT Faces Major Barriers	19
Post-Editing Will Dominate Translation Production in the Near Future	20
MT Will Enable Ever-Bigger Translation Volumes	21
What Does MT Mean for Europe?	22
Related Research	23
🔗 About Common Sense Advisory	24
🔗 Future Research	24
🔗 Applied Research and Advisory Services	24

Figures

Figure 1: Large LSPs Lead in PEMT Experience	6
Figure 2: Machine Translation Sees Major Gains by 2019	9
Figure 3: Enterprises Translate an Estimated 59% of Content with MT by 2019	9
Figure 4: The Machine Translation Adoption Curve	12
Figure 5: Customers, Employees, and Website Visitors Encounter MT	16
Figure 6: What Keeps Adopters from Increasing Their MT Investment?	20

Tables

Table 1: Demographics of Enterprise Respondents	3
Table 2: Demographics of Post-Edited MT Providers	4
Table 3: Enterprise Purchases of PEMT Focus on European Languages	8
Table 4: Three Classes of Enterprise MT Adopters	11
Table 5: Popularity of PEMT Production Models	14
Table 6: Methods for Translating Business Content Types	16
Table 7: Methods for Translating Customer Engagement Content Types	18

This page left intentionally blank

Executive Summary

Multilingualism is a core value of the European Union. In order to solve the integration challenges posed by 24 official languages, plus additional regional and minority languages, technology must play an important role. In a survey of 900 global enterprises, language service companies, and freelance translators about their experience with machine translation (MT), we found that Europe has taken a leading role in the development and implementation of this technology. Among our findings are the following:

- **MT provides a strong export market for European companies.** The majority of current demand for machine translation services comes from [North American tech firms](#), but the overwhelming majority of global supply comes from [small and medium enterprises in Europe](#), a sector that the European Commission has identified as “[the backbone of Europe's economy](#).”
- **MT is no longer just for big players.** In the past, machine translation was expensive, labor-intensive, and suitable [only for large enterprises](#). But recent developments – many of them in Europe and led by EU projects such as [Moses](#) – have made it [more accessible to the full spectrum of companies](#).
- **Post-editing leads growth in translation capacity.** Translation volumes continue to rise and enterprises target increased numbers of languages. “Pure” human translation [cannot meet anticipated volumes](#), but post-edited machine translation (PEMT) will [enable human translators to do more](#). Most of this increase in demand will be for [European languages](#).
- **Enterprises and LSPs have an expanding spectrum of options.** [Three enterprise MT adoption models](#) meet the demands of various sorts of buyers. At the same time, post-editing options are enabling [European LSPs to use PEMT](#) to improve productivity and meet client demand.
- **Business content leads, but user-generated content is increasing.** Core business content [leads commercial demand](#) for MT. User-generated content is challenging and [remains largely untranslated](#), but is especially important in the EU.

Overview of Enterprise MT

Machine translation – or MT – refers to software (an engine) that takes input in one language (the source) and automatically renders it in another (the target) without the need for human intervention. The direct output from such a system is raw MT. Optionally, a human may edit the output to improve it, a process known as post-editing.

Three approaches dominate MT today (see “[Rules-Based, Statistical, or Hybrid: Which MT Is Best?](#)” Oct11):

- **Rule-based MT (RbMT) systems rely on linguistic expertise.** They use complex linguistic rules and parsers to analyze language structures and map them to the target language. This approach – exemplified in systems such as [Lucy](#), [GramTrans](#), and the open-source [Apertium](#) – has been the subject of continual research since the 1940s.
- **Statistical MT (SMT) systems leverage big data.** They produce by comparing source content to a database of previous translations to find similar patterns. These engines, which emerged in the late 1990s, rely on access to large quantities of relevant bilingual texts. [Google Translate](#), [Bing Translator](#), and the open-source [Moses](#) system are among the most widely used translation technologies in the world. SMT is often less accurate – but more natural sounding – than RbMT.
- **Hybrid MT (HMT) systems deliver the best of both worlds.** Typically, an RbMT component makes a first pass and SMT “cleans up” the output. This approach takes advantage of rules where they work well but does not require the RbMT portion to cover all cases. Examples include [SYSTRAN](#), [PROMT](#), and [Asia Online](#).

The Role of MT in Europe

In order to compare how Europe engages with machine translation compared to the rest of the world, our survey covered a global audience. This approach allows us to examine differences between regions.

North America Leads in MT Demand

Consistent with previous research, we found that the bulk of enterprise demand for MT comes from North America, with Europe a distant second (see [Table 1](#)).

	Adopters		Non-Users	
Geographical distribution	Europe:	19%	Europe:	30%
	North America:	77%	North America:	68%
	Rest of World:	4%	Rest of World:	2%
Industry	IT:	63%	IT:	37%
	Services:	21%	Services:	37%
	Non-IT products:	16%	Non-IT products:	16%
Annual revenue	<€900 million:	23%	<€900 million:	58%
	≥€900 million:	77%	≥€900 million:	44%

Table 1: Demographics of Enterprise Respondents

Source: Common Sense Advisory, Inc.

Enterprise demand for machine translation skews heavily toward large corporations, both in Europe and elsewhere. However, we find that current growth rates – enabled by a burgeoning number of LSPs that use MT – are favoring smaller buyers of translation services, and CSA Research predicts that the majority of enterprises that market products internationally will adopt it in some form by 2020.

One of the reasons that North American enterprises lead demand for MT is that the IT sector has traditionally been its leading adopter and the largest technology companies are in the United States. For many years machine translation was an expensive technology that required considerable capital resources to produce and was economical only with high translation volumes. Our research finds that interest in the technology is broadening and shifting to new sectors, which will increase the role of European enterprises in the demand mix.

Europe Leads MT Production

In contrast to the demand side, Europe plays the leading role in the production of post-edited machine translation (PEMT) – the form of machine translation that enterprises are the most likely to use (see [Table 2](#)). Crucially, 81% of PEMT providers have fewer than 100 employees, a high adoption rate for an expensive technology among relatively small companies.

Region	Percentage of PEMT Providers	Percentage of Providers with Fewer than 100 Employees
Europe	56%	81%
North America	26%	69%
Rest of World	18%	76%

Table 2: Demographics of Post-Edited MT Providers

Source: Common Sense Advisory, Inc.

To some extent this finding reflects the fact that European LSPs have traditionally dominated the language industry and have also shown a willingness to embrace technology – much of it developed in EU-financed programs – that outpaces their rivals in the rest of the world. In addition, the dominance of U.S. technology firms on the demand side with strong European presence on the supply side reflects long-established “English outward” patterns in localization.

We found that nine of the top 15 countries for PEMT production are in the European Union: Spain, the United Kingdom, France, Germany, Italy, the Netherlands, the Czech Republic, Greece, and Portugal.

These patterns show that MT – like translation in general – is largely an export market for European firms. They sell services to companies in other countries – primarily the United States – that seek to do business in Europe.

At current growth rates, PEMT will be a leading driver for growth in the European language services sector in the coming years. Our research suggests that enterprises intend to increase their translation volumes by 67% over current levels by 2020. They will concentrate most of this growth in PEMT, which has a 36% compound annual growth rate. By contrast, unaided human translation will see only modest growth rates of around 4% per annum.

Drivers for MT Adoption

This section considers the business factors that lead enterprises to work with machine translation technology, what they hope to gain from it, and how they deploy it within a broader global content strategy. It covers three areas:

- 1) [enterprise size and experience with translation that lead them to adopt MT](#);
- 2) the [business value of translated content and how well MT meets buyer demand](#); and
- 3) [the content volumes they deal with](#).

Large Enterprises and LSPs Led the Way

As noted above, MT use levels correspond strongly with enterprise size: The larger an organization is, the more likely it is to employ machine translation as a content localization tool. In particular, large corporations are far more likely than smaller ones to develop their own engines. This finding is not surprising: Installing and maintaining MT systems is expensive and resource intensive. At the same time, translation is not a core business activity for most of them, so it makes sense to outsource it as a task. Only when faced with high volumes or the need for quick turnaround time does it make sense to bring this task in-house.

For similar reasons, in the past, few LSPs could successfully muster the resources needed to build their own systems. MT developers such as [Lucy Software](#) and [Systran International](#) – both based in Europe – offered systems that they could implement, but the cost of customization made these systems practical primarily for providers that serviced large clients with high volumes. Most LSPs found machine translation impractical due to their diverse client bases and smaller volumes. Over the years Europe has led in MT research and development, only to see much of its talent bought up by American tech firms.

In our research we find that size correlates with MT experience. Large LSPs generally have more experience with using it in their workflows than do smaller ones (see [Figure 1](#)). At the same time, small LSPs with 20 or fewer employees actually outnumber other providers at all experience levels up to 10 years: They are the most numerous group in the industry, and even low adoption rates among them yield greater numbers than very high percentages among large translation companies.

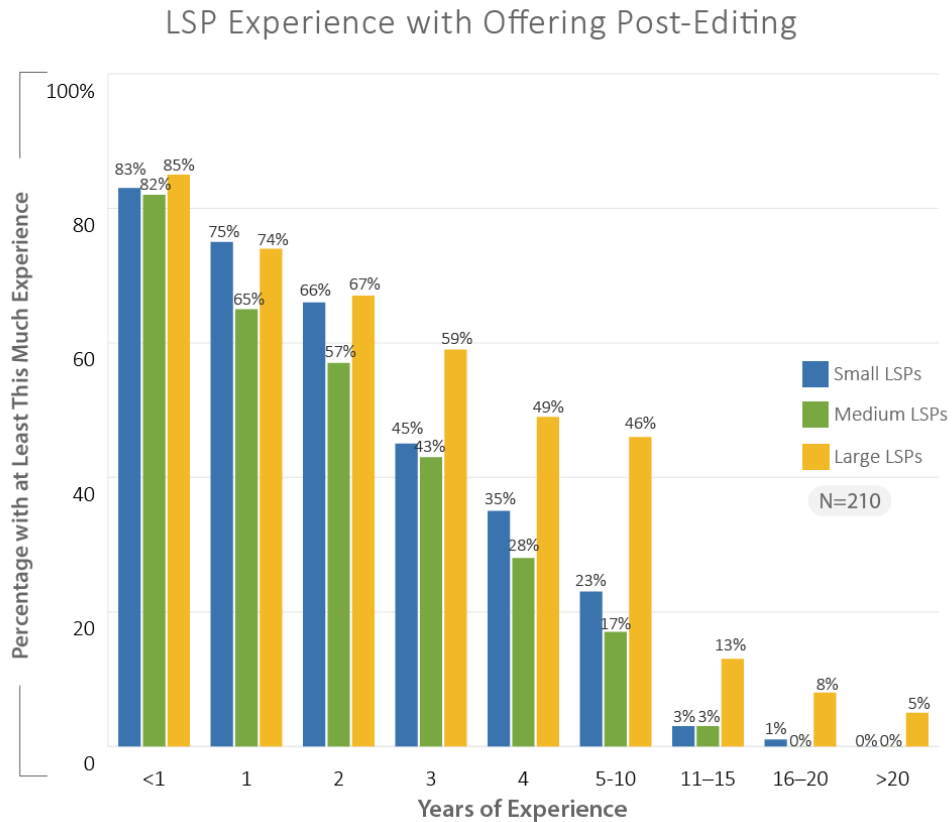


Figure 1: Large LSPs Lead in PEMT Experience

Source: Common Sense Advisory, Inc.

Smaller Ones Are Catching Up

Even though large providers and consumers were historical leaders in machine translation, today the mix is shifting toward smaller LSPs and enterprises.

What changed? One of the biggest factors was the release of [Moses](#) version 1.0 in 2013. Developed through the EU-funded [MosesCore project](#), this open-source software made state-of-the-art statistical MT accessible to any enterprise of LSP that has the needed technical infrastructure and can find or develop the needed skill sets. Many organizations that had avoided the technology in the past due to licensing fees or data security concerns – associated with sending their content to tech giants such as [Google](#) and [Microsoft](#) – saw Moses as an attractive alternative.

In addition, a new breed of hosted MT providers such as [CrossLang](#), [Kantan](#), and [Tilde MT](#) – many of them providing custom Moses installations – has simplified access for small language companies and enabled them to use it. Many have flocked to it since that time.

Roughly one in three LSPs that offer post-editing services worldwide is a small European provider with 20 or fewer employees, a number that rises to almost half when considering those with fewer than 100 employees. These relatively modest companies are the single largest block of implementers by a significant margin. Turning to European LSPs that do not currently offer PEMT services, we find that they are almost three times as likely to seriously consider adding them to their portfolio compared to LSPs elsewhere in the world.

Because MT utilization rates are already quite high for large and mid-sized LSPs – which constitute a pool that will not expand dramatically – the real growth in PEMT adoption and corresponding opportunity for technology providers is among small LSPs – particularly in Europe.

MT Gains Traction for High Volumes in European Languages

Enterprises employ MT for a relatively small subset of their languages, but they use it to produce large quantities of content, taking advantage of its speed and low cost (see [“Transformative Translation,”](#) Oct13). Broadly speaking, demand for PEMT follows that for languages in general (see [“Digital Opportunity: Top 100 Online Languages for 2016,”](#) Apr16). However, after the top 20 languages, it falls off rapidly. An LSP that specializes in German is much more likely to have its clients request PEMT than one that translates into Hindi or Hungarian.

Looking more closely at the top languages into which enterprises translate their content, we find four of the five languages (and 11 of the top 20) with the most demand for PEMT are official in the European Union (see [Table 3](#)). Three of the remaining languages – Russian, Norwegian, and Turkish – are not official in the EU but nevertheless play an important role within Europe. Other leaders include East Asian tongues: Given the importance of trade between the EU and Asia, expertise in them – or partnership with LSPs/MT providers in countries where they are spoken – can provide a strategic benefit for European language companies.

Language	Percentage purchasing PEMT (N = 83)
French	64%
Spanish	61%
German	60%
Portuguese	57%
Simplified Chinese	55%
Italian	53%
Japanese	52%

Language	Percentage purchasing PEMT (N = 83)
Russian	51%
Traditional Chinese	42%
Korean	41%
English	34%
Dutch	31%
Swedish	30%
Polish	29%
Danish	27%
Norwegian	25%
Finnish	24%
Turkish	23%
Arabic	23%
Hebrew	17%

Table 3: Enterprise Purchases of PEMT Focus on European Languages

Source: Common Sense Advisory, Inc.

Note: Official EU languages are in bold face.

Increasing Translation Volumes Make MT Inevitable

We asked enterprise respondents about their plans for how much content they will process using different translation modalities. We found two different stories, depending on whether we look at the median values – which tell us what typical companies do – or averages – which provide a picture of what the overall market will do. Medians show us that most enterprises work with relatively modest amounts of source content and that human translation currently leads as the most popular method for production of localized content (see [Figure 2](#)). Averages show much higher volumes – reflecting a few respondents with very high volumes – and smaller percentage increases.

One of the most important observations we see is that the typical enterprise is going to reduce its investment in human translation, but will almost exactly offset this change with an increase in the use of PEMT. As a result, LSPs will see many clients shift to MT-centric workflows. These shifts seemingly confirm widespread fears that technology is destroying translators' jobs. However, at the macro level we will see some buyers increase their reliance on human translation substantially. The results mean that even as machine translation appears to erode the market for professional linguists' services, we are actually seeing modest increases with substantial growth in PEMT.

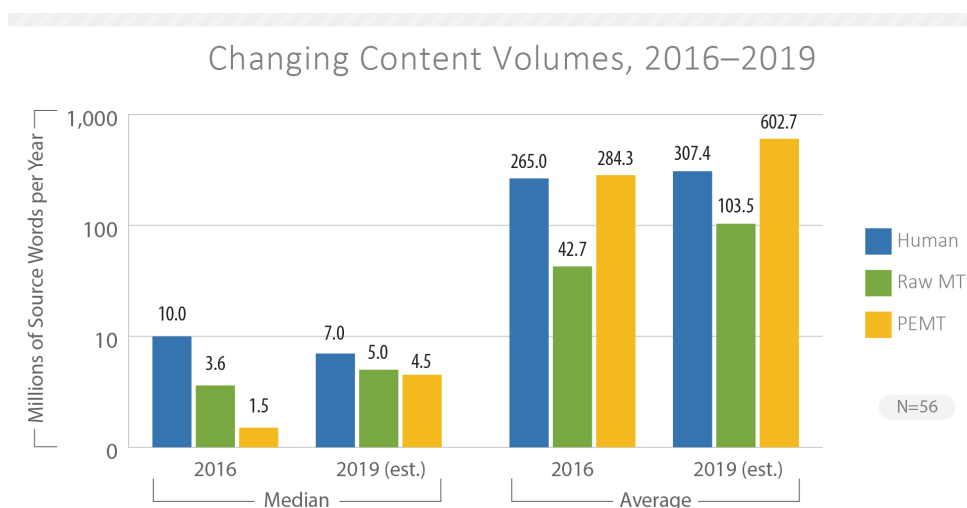


Figure 2: Machine Translation Sees Major Gains by 2019

Source: Common Sense Advisory, Inc.

In other words, the pie is getting bigger, even though how it is sliced is changing (see [Figure 3](#)). LSPs that are willing to work with post-editing will see increases in demand for their work while those that avoid it may see stagnant or even negative growth. European LSPs – with their relatively tech-friendly outlook – stand to have a competitive benefit in this regard compared to other regions.

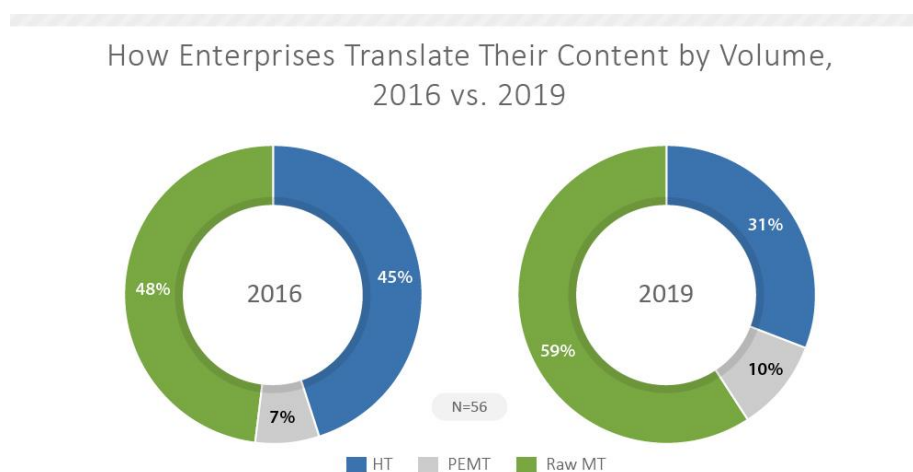


Figure 3: Enterprises Translate an Estimated 59% of Content with MT by 2019

Source: Common Sense Advisory, Inc.

MT Adoption Models

Looking at how enterprises and LSPs adopt machine translation, we found various models. These models correlate with organizational size and resources, as well as with what the adopters hope to gain. In this section we briefly explore these models and examine how European companies compare to the rest of the world in using them.

Enterprises Have Three Models for MT Adoption

In this section we define three models for enterprise MT adoption.

Three Enterprise Models

In our research we found three distinct models for how enterprises adopt MT based on where they produce the MT they work with (see [Table 4](#)). European enterprises show a slight tendency to bring MT production in-house compared to the rest of the world. We define these three groups as follows:

- **“Toe Dippers.”** The fastest growing group of MT implementers, they are risk-averse and look for low-cost approaches to machine translation that minimally disrupt their existing content strategies. They primarily outsource both MT production and post-editing, and treat PEMT as a drop-in replacement for “pure” human translation. They look for bargains and ease of use. They have the least experience with MT and expect little from it, other than better speed, lower price, and increased volume.
- **“Content Busters.”** This group produces large volumes of content and brings machine translation in-house to support their requirements. They translate too much content and want too much control to leave production to third parties. They translate into fewer languages than other groups, but are strongly oriented toward growth in volume and number of languages. They have considerable experience with MT and translation and a good idea of its pros and cons.
- **“Turnaround Artists.”** These organizations are primarily concerned with decreasing turnaround time and maintaining language coverage, so they adopt a mixture of internal and external production that allows them the greatest flexibility in MT production. They do not have particularly high volumes, but they are very skilled in translation-related activities and understand what they can expect from technology. They translate into more languages than others.

Characteristic	Toe-Dippers	Content Busters	Turnaround Artists
Production strategy	Outsource everything	Generate in-house	Pragmatically mix internal and external models to deliver faster
Percentage of adopters	Europe: 45% Rest of World: 55%	Europe: 35% Rest of World: 20%	Europe: 20% Rest of World: 25%
Annual revenue	Any size	>€10 billion	>€10 billion
Typical source volume	<15 million words	>15 million words	<15 million words
Typical number of MT target languages	Up to 10	Up to 20	40 or more
Drivers for MT adoption	Cost savings	Content volume	Number of languages and speed
MT experience	Low	High	Highest
Growth as a group in next three years	High	Low	Low

Table 4: Three Classes of Enterprise MT Adopters

Source: Common Sense Advisory, Inc.

MT Adoption Follows a Curve

The tension between investment and risk on the one hand and the benefit of increased throughput and speed on the other leads to a distinct pattern in how organizations adopt MT over time (see [Figure 4](#)). Historically, we find that enterprises:

- **Start with human translation.** Their global expansion starts with outsourced human translation. Those with modest needs or that work in fields with stringent quality needs stay here.
- **Take baby steps with MT.** After they build experience with human translation, and if their volume reaches levels where it is too expensive or time consuming, they start MT pilot tests for a few core languages or projects.
- **Adopt outsourced MT as a core translation strategy.** After gaining experience on a small scale, they systematically use PEMT for technical and structured documentation. It does not replace human translation – which they retain for marketing and non-technical materials – but instead adds to it. For most enterprises, this is the final step.
- **Bring the technology in-house.** Organizations with sufficient content volume or a need for faster throughput may eventually internalize some

or all of their MT production, using a combination of raw and post-edited output. In general, only the largest enterprises – those with revenue greater than roughly €1 billion – can take this step.

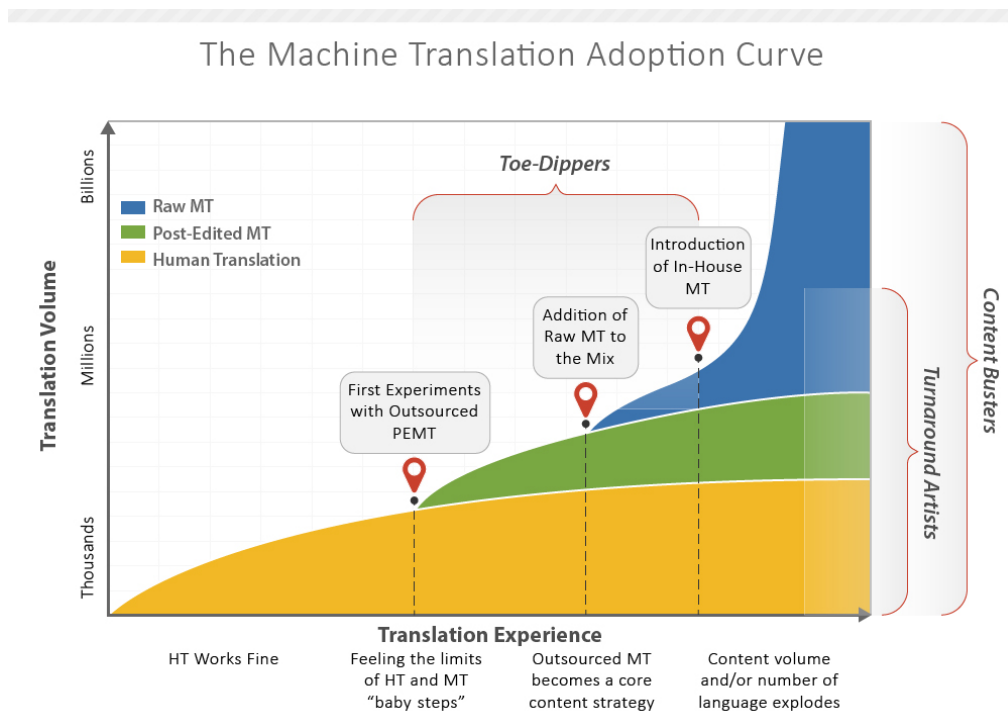


Figure 4: The Machine Translation Adoption Curve

Source: Common Sense Advisory, Inc.

The Curve Is Changing

The curve above describes a typical enterprise journey with MT. It is, however, changing. The increase in the number of providers of post-editing and its increasing acceptance as a mainstream production method now mean that many enterprises can accelerate the first steps and go directly to using machine translation as a core content strategy. Because PEMT is a drop-in replacement for human translation, it makes sense to move to MT as soon as an enterprise has a sufficient body of human translation to serve as training data for an MT engine. This can cut the time to first MT implementation by several years.

LSPs Move to Adopt MT

Language service providers – also known as “language service companies” in the EU – face a fundamental choice when it comes to machine translation: Do they avoid it as a threat, or do they embrace it as an opportunity? Historically LSPs resisted MT, but increasing numbers – particularly in Europe – are coming to see it as a strategic opportunity and necessity.

In our research we found five models of engagement with PEMT. Note that these models are not mutually exclusive, and one third of LSPs combine them, a number that surely underestimates the actual levels because so many translators use MT as a productivity tool without telling their clients.

1. **Non-users.** Roughly 35% of our respondents do not offer post-editing services at all. Another 16% are preparing to but are not ready to roll them out on live projects yet. Their engagement is casual or opportunistic: They may occasionally use online MT for term lookup or other casual needs, but do not employ it systematically as a business tool. However, it is likely that linguists in their supply chains already use MT to improve productivity.
2. **Free MT as a productivity tool.** Thanks to modern computer-assisted translation (CAT) tools, most linguists have access to results from one or more free engines alongside long-time features like translation memory and terminology lookup. Our research shows that many freelancers use these features, even if they do not inform their clients.
3. **Post-editing as a service.** LSPs in this group do not have their own MT production capacity and receive raw output to edit directly from their clients or other LSPs. This approach requires little or no infrastructure investment, but does require expertise in post-editing content.
4. **Shallow service.** LSPs work with generic online engines – either using free services or via paid APIs such as the [Google Translate API](#) – to generate the output they edit. They save some money because they do not invest in MT technology or a subscription to a cloud service. However, they typically lack real integration between the engine and other tools and have no way to improve the process or quality of the output.
5. **Integrated strategic solution.** LSPs use trained MT systems – either their own or dedicated systems managed by a third party – that produce raw output that they then edit. Regardless of where their engines are, they manage the training on behalf of their clients. They almost universally integrate the engines with their translation memory systems to combine the benefits. They often provide their machine translation via their CAT tool to allow their linguists to work with it in a familiar environment.

Of these models, the integrated strategic solution and post-editing lead the list and are roughly equal in popularity (see [Table 5](#)), with free MT as a productivity tool in third place, and the shallow service model in last place. In terms of the models they adopt, European LSPs look much like those in the rest of the world, except that they are slightly more likely to implement multiple models and are more likely to use free MT as a productivity tool.

LSP Location	Service Model			
	Free MT as a Productivity Tool	Post-Editing as a Service	Shallow Service	Integrated Strategic Solution
Europe	25%	50%	16%	56%
Rest of World	16%	58%	14%	56%

Table 5: Popularity of PEMT Production Models

Source: Common Sense Advisory, Inc.

Audiences and Content Types

Enterprises deploy MT to improve the customer experience (CX). Even though other audiences are important, it is the need to engage and retain customers in international markets that drives their adoption. However, enterprises are selective in when, how, and for whom they apply it. This section examines the ways in which adopters choose and address their audiences to help bring international CX up to par with the level they deliver for domestic audiences.

Enterprises Aim MT Output at Underserved Customers

Organizations employing MT – either raw or post-edited – understand and respond to the need to communicate with their customers in their languages (see [“Benchmarking the Top 100 Online Languages for 2015,”](#) Apr15, and [“The Rise and Fall of the Top Online Languages,”](#) Apr15). They use MT pragmatically to extend their reach and stretch translation budgets (see [“Finding Revenue in Under- and Over-Served Languages,”](#) Sep15).

We found that the majority of enterprise MT adopters target three audiences with their translation efforts: customers (93%), website visitors (70%), and business partners (51%). Fewer than half aimed for the remaining three categories: employees, prospects, and search engines. There is no measurable difference between European enterprises and others in this area.

However, the targeted demographics for MT differ from the general ones. After removing those respondents who were unsure from the total, we calculated the percentage that expose MT to each audience. We then multiplied the result by the percentage of respondents that targeted each audience overall. These results show how likely each one is to actually encounter MT (see [Figure 5](#)). Due to the low numbers for each category, we do not break these figures down by production or post-editing method.

According to this analysis, the customer remains the most likely audience to see MT, but employees rise to the second position, with website visitors and partners close behind. Prospects are unlikely to see MT, as are search engines – even though MT adopters are more likely to target them. These two areas are ones where enterprises often prefer a hands-on approach because they relate directly to success in sales.

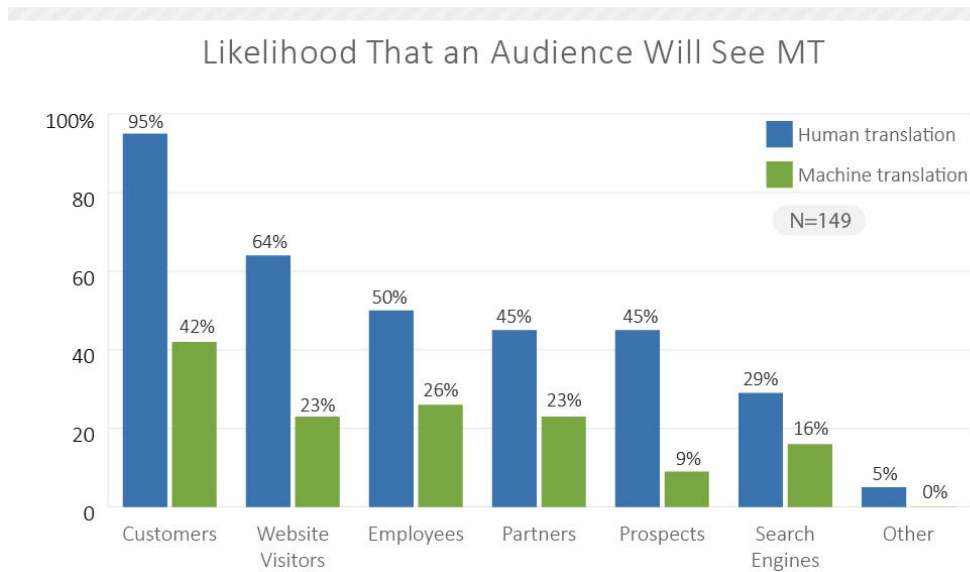


Figure 5: Customers, Employees, and Website Visitors Encounter MT

Source: Common Sense Advisory, Inc.

Business Content Leads Machine Translation

MT adopters are selective in how they use the technology. They focus on business rather than user-experience content. They prioritize core types that they generate in-house and slowly add other types when they experience the need. We found that global MT deployment rates exceed HT rates for four of the 10 types of business content we asked respondents about (see [Table 6](#)).

Content Type	% Translating (Global / Europe)	Translation Method (N = 83)	
		HT (Global / Europe)	MT (Global / Europe)
Marketing	95% / 100%	90% / 88%	25% / 35%
Product documentation	94% / 88%	59% / 59%	66% / 59%
Websites	94% / 100%	81% / 88%	53% / 53%
Online help	93% / 88%	63% / 59%	65% / 53%
Training materials	89% / 94%	73% / 71%	43% / 29%
FAQs	81% / 88%	51% / 65%	57% / 47%
Business forms	80% / 100%	69% / 88%	27% / 29%
Catalogs and e-commerce	73% / 82%	63% / 76%	30% / 29%
Knowledge bases	72% / 82%	48% / 71%	49% / 41%
Support e-mails	71% / 82%	53% / 71%	36% / 41%

Table 6: Methods for Translating Business Content Types

Source: Common Sense Advisory, Inc.

Note: Bolded rows show where MT rates exceed those for human translation.

In addition, we found that:

- **European enterprises use MT in the same ways as others.** As seen in [Table 6](#), European respondents adopt machine translation at levels similar to those elsewhere, and for the same types of material. The differences seen are not significant given the small number of responses from European enterprises to this question.
- **Marketing materials and business forms do not mix with MT.** Although marketing material is the most commonly translated business content (95%), only 25% of respondents touch it with any form of MT. When they do, it is usually only after post-editing. Business forms, catalogs, and e-commerce content show similarly low rates even though technical catalogs are an ideal application for MT.
- **Machine translation outscores human for four content types.** Product documentation, online help, FAQs, and knowledge bases (highlighted in [Table 6](#)) all have MT usage either statistically indistinguishable from or higher than HT rates.
- **Generic MT is not a serious choice for most business content.** Respondents seldom use free online tools such as [Google Translate](#) or [Yandex.Translate](#) for any business content. It cannot deliver the organization-specific terminology or language that they need, even for “low-quality” usage scenarios. As a result, they prefer to work with trained MT when possible. However, they may use untrained as input for PEMT production if they do not have access to trained systems and turn to it informally to support internal needs.

Customer Engagement Content Remains Largely Untranslated

We also queried our respondents about their translation strategies for user-generated content (UGC) that customers, visitors, and partners create – including blogs and comments, group discussions, chat and messaging in various forms, tweets, forums, and user reviews – and similar material that they generate internally. As in our previous research, we found that much of this goes untranslated: Most respondents (64%) do not translate any of this material (see [Table 7](#) and “[Transformative Translation](#),” Oct13). Those that do are, on average, 1.67 times *more* likely to use MT than HT for it.

For this material, MT plays an important role. UGC has limited or uncertain value, and much of it would remain unread if enterprises were to translate it. At the same time, some of it can be extremely valuable, but within a short window. For example, if a chat message goes unanswered for more than 60 seconds due to translation requirements, the person who sent it likely will leave unsatisfied.

Content Type	% Translating	Translation Method (N = 83)	
		Human (HT)	Machine (MT)
Blogs and comments	37%	22%	22%
Forums	31%	12%	25%
User reviews	29%	16%	23%
Chat, instant, or SMS messages	27%	8%	24%
Group discussions	25%	11%	19%
Tweets	20%	12%	10%

Table 7: Methods for Translating Customer Engagement Content Types

Source: Common Sense Advisory, Inc.

Note: Bolded rows show where MT rates exceed those for human translation.

These results show that little funding for translating non-core content – such as user-generated and interpersonal material – is generally available, although supplemental content may appear on the radar. If enterprises do not see the value, they are unlikely to invest in it. However, if they do translate it, the uncertain value and short shelf life argue for the use of raw MT – despite its difficulty and limitations – rather than human translation or PEMT, both of which may take too long or be too expensive. Enterprises often prefer an on-demand, automatic approach because they know that the overwhelming bulk of customer engagement content will remain unread in most languages.



Within the European Union, the focus on e-Inclusion and e-Citizenship makes these content types more valuable for the public sector than they typically are for businesses. As a result, Europe needs to find ways to deal with UCG in a cost-effective manner with good-enough quality to meet user needs. Developing a system that meets these requirements and that allows speakers of Europe's official and major regional languages to communicate effectively on a personal level remains a major challenge.

MT Faces Major Barriers

Enterprises today use MT for a small portion of their languages. Most anticipate that they will increase their use, but they don't expect MT to replace humans en masse any time soon. Without post-editing, the technology remains unsuitable for most core content, so professional linguists will see demand increase for both "classic" HT and post-editing. Although they would be foolish to ignore machine translation, their jobs will remain secure – albeit altered – for now.

MT offers compelling advantages in cost, speed, and throughput. If you want to do more with MT, what challenges are you likely to face? Regardless of how you deploy MT, consider the following factors (see [Figure 6](#)):

- **Quality is a stubborn problem.** Three-quarters of our respondents see quality as a major barrier to further deployment of MT (see "[The Quality-Availability Debate around MT](#)," Oct13). You are likely to find that post-editing is necessary – at least for core content types.
- **Technical complexity and integration challenge even savvy adopters.** Your authoring and publication environments may not play nicely with MT. Many adopters report that integration with translation memory tools poses a notable challenge, even with APIs. However, if you have the experience, volume, and resources to bring everything in-house, you may find integration to be less of a challenge after you get past the initial investment.
- **Formatting is the unsolved problem.** MT engines deal well with plain text. Throw in formatting codes or other tags, and they break down. System developers try to engineer around these problems, but for each new format you add, you may find that you have to repeat the process.
- **Qualified staff is hard to find.** If you wish to bring MT in-house, you may find it difficult to recruit staff with the requisite skills. MT is still a young field, and the market for individuals with a strong track record for deployment is competitive.
- **Concerns about data security are crucial for some adopters.** If you work with financial or personal data, free online MT can be a big problem because you risk revealing confidential information (see "[Data Leakage from Free Machine Translation](#)," Nov13). These issues are especially important in Europe due to the European Data Protection Directive.

- **Training engines isn't simple.** It isn't enough just to hand your translation memories over to an MT supplier. To realize the full benefits of machine translation, you need an ongoing data curation process for removing outdated or problematic materials. You must also manage your terminology and implement processes to ensure continuous improvement.

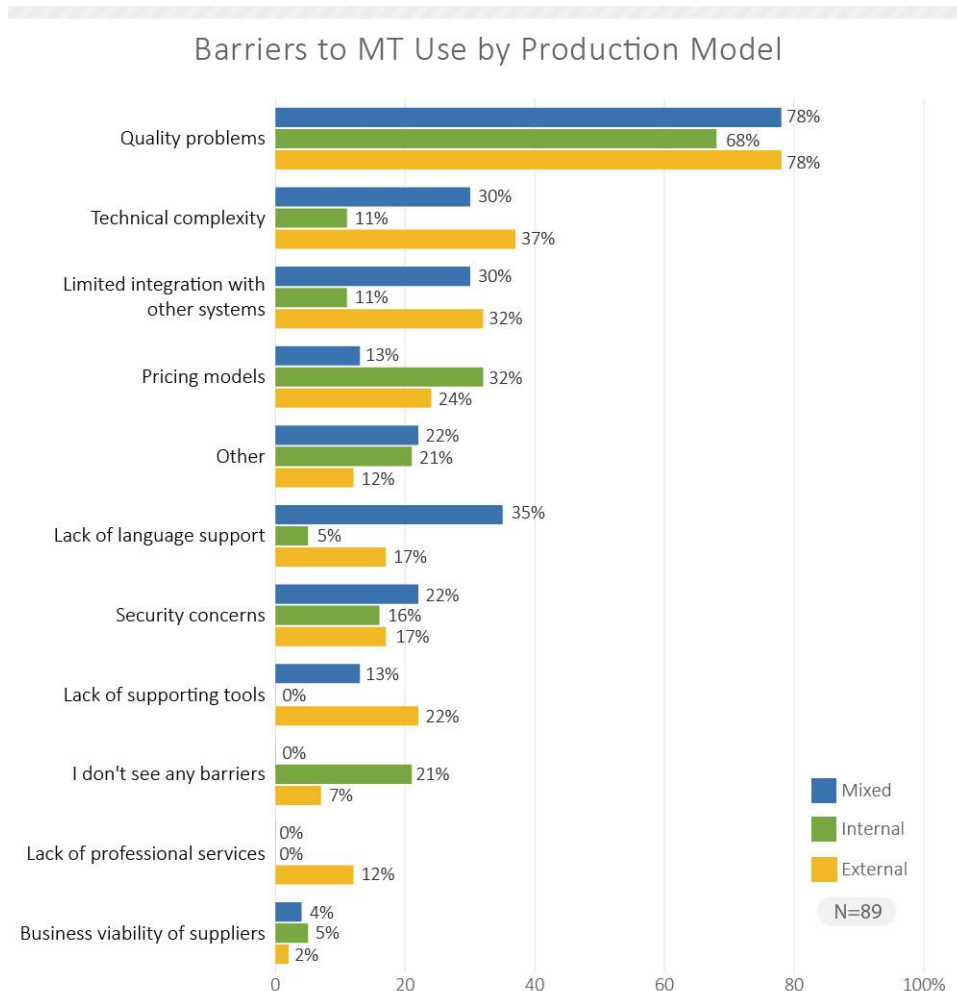


Figure 6: What Keeps Adopters from Increasing Their MT Investment?

Source: Common Sense Advisory, Inc.

Post-Editing Will Dominate Translation Production in the Near Future

Most professional translators have negative opinions of post-editing MT: In our 2016 survey of freelance translators and their use of MT, they assigned it an average of 3.7 on a scale of 1 ("I hate it") to 10 ("I love it"), with 30% giving it the lowest possible score and only 17% assigning a truly positive rating. However, they will not be able to avoid it. Their customers – and the LSPs that outsource to them – will increasingly demand post-editing services. PEMT is the wave of the future, at least for high-value content. Depending on

the type of post-editing, per-person throughput can be two to eight times what a linguist alone can achieve (see "[Post-Edited Machine Translation Defined](#)," Apr13).

For buyers, the decision to use PEMT is a no-brainer. It offers them equivalent translation output at a fraction of the cost and in less time. It combines with other technologies – such as translation memory and workflow management tools – to enable them to translate more content or support additional languages that extend their global reach (see "[Benchmarking the Top 100 Online Languages for 2015](#)," Apr15).

We find a wide range of pricing for PEMT (see "[What Post-Edited Machine Translation Costs](#)," Mar13). However, prices for "heavy" post-editing are slowly stabilizing at around 65% of the price for full human translation. Trained MT based on and integrated with an enterprise's translation memories can bring costs down even more dramatically by combining the savings possible through each technology. Cost-driven adopters can find lower rates – around 45% of the cost of HT – but often settle for a lower level of post-editing to obtain them.

MT Will Enable Ever-Bigger Translation Volumes

Our respondents reported that they intend to increase translation volumes by 67% over the next three years, from an average of 590 to 990 million words per year. Even though these numbers are much higher than the median that most enterprises translate, they point toward an increasingly common scenario for large organizations.

Our respondents cannot increase their volumes with human translation alone. The growth they forecast would exceed the capacity of all current translators, as well as those who plan to enter the field in the foreseeable future (see "[The Calculus of Global Content](#)," May16). No alternative to MT can meet the needs of enterprises – and those that do not adopt some form of it will find themselves left behind.

What Does MT Mean for Europe?

Based on our research, we close with observations and recommendations for MT in Europe:

1. **Europe may lead development and adoption, but capitalization lags.** Many of the most important MT advances in recent decades have come from Europe and EU-funded projects. Nevertheless, the biggest developers are U.S.-based tech firms (such as [Facebook](#), [Google](#), and [Microsoft](#)) that have staffed their research programs with European participants or bought European technology. If Europe is to remain competitive, it will need to find ways to capitalize public investment at home.
2. **Machine translation provides an economic opportunity for Europe.** We have found that MT-centric language service providers have annual growth rates almost 3.5 times those of competitors that are more conservative in their approach to the technology. With so many LSPs based in the European Union, the technology can drive substantial growth among small and medium language companies.
3. **Open-source projects lead the way.** The release of Moses based on EU-funded work marked a watershed moment in MT: For the first time, implementers could use a shared technology stack accessible to even relatively small companies. Many current MT providers have built their programs on the basis of Moses. As newer technologies emerge from research projects, it will be important that the results make their way into similarly open and accessible outcomes.
4. **More research is needed on extending MT to user-generated content.** Machine translation does well for technical documentation, especially when post-edited. However, many of the areas where MT can offer the most social benefit require systems to deal with very diverse language without editing. These areas are particularly challenging for current-generation MT. Newer technologies that may help are still under research and require more funding and field testing.
5. **Europe's lead in this field can benefit society.** European expertise in MT can help reduce language barriers, but the focus has to shift from basic research to implementation with defined outcomes and benefits. The European Commission can take a leading role in making this happen.

Related Research

CSA Research has been studying and analyzing the market for machine translation since 2004. We recommend the following CSA Research reports and briefs on related topics (accessible to CSA members):

- [“Transformative Translation”](#) (Oct13) – This report analyzes the conflict between content quality and availability – and the role that machine translation (MT) plays. It outlines the transformation that accompanies the decision to make information available when needed.
- [“Human-Enhanced Machine Translation”](#) (May13) – This report reviews the experiences of organizations that buy post-edited MT services from external suppliers. It references [“Post-Edited Machine Translation Defined”](#) (Apr13), which defines the most common types of post-editing, characterizes the marketplace for PEMT, and provides several examples of the process.
- [“Content Strategy for the Global Enterprise”](#) (Apr11) – This report describes the content challenges faced by most enterprises, provides a typology of content, and recommends where machine translation fits best.
- [“Trends in Machine Translation”](#) (Oct11) – This report analyzes the MT market, the technology, and the corporate and business factors that are guiding the evolution of this technology.

In addition, the following publications from EU-funded projects provide insight into the European language technology landscape and the role of MT in Europe:

- [“META-NET Strategic Research Agenda for Multilingual Europe 2020 \(SRA\)”](#) – The SRA raises awareness of Europe's language technology industry and how it relates to EU development priorities.
- [“Strategic Agenda for the Multilingual Digital Single Market”](#) – This document showcases a vision in which language technology works to overcome “language blocking” to support greater participation and integration within Europe.
- [META-Net White Papers Series](#) – The 32 volumes of this series provide an overview of European languages, digital support for them, and their outlook in a digital age.

About Common Sense Advisory

Common Sense Advisory, Inc. is an independent research firm committed to objective research and analysis of the business practices, services, and technology for translation, localization, and interpreting. With its research for both [Global Leaders](#) and [Industry Providers](#), Common Sense Advisory endeavors to improve the quality and practice of international business, and the efficiency of the online and offline operations that support it. To find out more about our research and how to become a member:

- E-mail us info@commonsenseadvisory.com.
- Visit www.commonsenseadvisory.com.
- Call +1.978.275.0500.

Future Research

Common Sense Advisory seeks interviewees from the community of people involved in building business applications for international use. If you would like to be interviewed or have clients who would like to share their experiences, please e-mail us at info@commonsenseadvisory.com. We anonymize participants and hold all information in the strictest confidence.

Applied Research and Advisory Services

This report and other Common Sense Advisory research into the best practices of business globalization serve as the foundation for our Applied Research and Advisory Services including International Customer Experience Assessments, Vendor Selection, Localization Business Process Audits, Globalization Excellence and Optimization Assessments, and Globalization Roadmaps. E-mail us at info@commonsenseadvisory.com for more information.

