Let’s run the world.

There’s more to running than miles.

Running is about the distance the run carries you. It’s about the terrains you cross, the trails you take, and the roads less travelled—for every high, a low; for every stretch, a bend, and for every finish, a start.

At Smartling, we believe in the power of words, and more importantly where words can take you. The way we see it, words are just words until they are given meaning and context. That’s the beauty of language and the power of translation.

The world runs on words. You should, too.

Visit Smartling.com/why-smartling to learn why we’re the #1 rated TMS by users on G2.
We believe that humans make the difference. We believe that language is at the core of what makes us human. That technology helps us to perform at our best and increase our capacity. The better we perform, the more we can invest in people. And the more we invest in making people happy, the better we perform. We support talent through education, we invest in young people’s ideas, we trust our clients. We are building the most engaged community of translators ever. A virtuous, long-term symbiosis where humans are supported by machines and machines learn from humans. This allows us to be efficient, adaptable, convenient and fast.

We open up language to everyone.
We found ourselves in the middle of a global pandemic halfway through putting this issue together. And we’ve decided with the advent of the novel coronavirus sweeping the planet, we are suspending print issues for now. Physical mail delivery is difficult to certain parts of the globe in the best of times, and a pandemic will certainly not make it easier. The upside is, we’ll save paper in a time when paper seems to be in high demand — reducing our carbon footprint and making the demand for wood pulp just a little less.

Of course, it felt bittersweet, going over the pages knowing they might be the last to make it to physical prints. I’ve been with the magazine for 12 years, and this is a first for us.

The digital issues go on, however. Given that our issue on AI and machine learning is full of new information and data — from machine translation to chatbots to AI in project management — this seems to be the way the entire world is going. Once you have the infrastructure in place, digital is smart on every level. It’s near-instant, and helps keep social distancing in place when germs are a concern.

So please enjoy this issue, and look for future issues in your email inbox!
SDL Trados GroupShare 2020

Business intelligence at your fingertips
Combining advanced online translation editing with insightful reporting and new project management capabilities, SDL Trados GroupShare 2020 offers you a new, smarter way to create, translate and measure your work.

New to GroupShare 2020:

- Advanced, online translation editor
- Enhanced reporting and analytics capabilities
- Customizable and exportable reporting
- Greater online project management functionality

For more information, or to begin your free trial, visit sdl.com/groupshare2020
Focus: Artificial Intelligence & Machine Learning

24  AI in the translation business: Beyond MT
    Dmitry Ulanov

28  How AI is transforming the content supply chain... and your job
    Jessica Roland

33  Multi-engine MT management for global companies
    Marco Zappatore & Filip Šanca

38  Toward global machine translation
    Arvi Hurskainen

44  Modern tech meets Indian Sign Language
    Ribhu

48  When internationalization isn’t enough
    Arle Lommel

52  How multilingual chatbots influence localization
    Kaspars Kaulins
About the cover:
Synchronized robotic devices with monitor screens create a choreography of intelligent movement at the Museum of Science and Industry in Chicago.
Would you introduce yourself?
Wafaa Mohiy, general manager of the Saudisoft Co. Ltd. Egypt office.

Where do you live?
Cairo, Egypt.

How did you get started in this industry?
It has been a long journey. I graduated with a computer science major and worked as a business application software developer for the first three years of my career. It was a good start, but I did not find my passion in developing business applications. I took systems engineering courses and afterwards I specialized in enabling applications/OS to support Arabic languages, as there were not many back then!

In the early 1990s, I participated in enabling the first Arabic GUI operating system on a PC, which was based on an English OS called GEM. I really found my passion enabling different applications to support Arabic and complex bidirectional features. Then, I moved to Saudisoft as a software development manager, and my first project was developing Arabic file converters for Microsoft Word 2.0. Afterwards, we were asked by Microsoft to start our first localization project (Microsoft Access 2.0). It was a huge localization project with no CAT tools, no industry standards or even experience. Despite the difficulties, we successfully delivered the project, and since then I have been involved in the localization industry, which I found very interesting and informative. It also helped me to develop (professionally and personally) through the years.

How long have you worked in it?
More than 25 years now!

What language(s) do you speak?
Arabic as my mother tongue, and I’m fluent in English. I also studied French at school and took German courses as I really love learning new languages. Unfortunately I could not master any other language like English, because I don’t practice them.

Whose industry social feeds (Twitter, blog, LinkedIn, Facebook) do you follow? (If any)
I follow LinkedIn on a daily basis to keep updated with industry news, events and reports. I am also connected with many colleagues in the industry. Besides localization, I also follow many technology providers groups to keep updated with the latest trends and innovations.

What do you like to do in your spare time?
When I have some free time, I either read or have coffee in a place with a water view. I also like traveling to coastal cities for long weekends.

What industry organizations and activities do you participate in?
After being in the industry for so many years, there was this moment where it was important to get involved in different activities and associations. Paying back into the industry and community is really rewarding and it is a real use for the experience gained. It is a way to share knowledge and gain new expertise. Some of the activities I have participated in were as program committee member for LocWorld and as a GALA (Globalization and Localization Association) board member for four years (January 2016 to December 2019). I was also a speaker and program committee member at the American University in Cairo with the first translation and localization conference in Egypt in 2018. With other local industry professionals, we were able to establish EAGLS (Egyptian Association for Globalization and Language Solutions), the first localization association for the middle east region, and we are running many events for industry professionals locally.

Why do you read MultiLingual?
MultiLingual has been always the place to go to for industry updates, trends and events. It covers topics from every angle: technical, linguistic, tools, processes, events and so on.
People

Recent industry hires

- Lionbridge, a provider of localization services, has announced that Corinne Saunders has been hired as managing director, EMEA. Lionbridge www.lionbridge.com
- XTM International, developers of a translation management system and computer-aided translation tool, has hired Justin Bechtel as business development manager, Monika Voorhis as solutions architect, Sara Basile as product manager and Dave Ruane as digital content and partnership manager. XTM International https://xtm.cloud
- Iconic Translation Machines Ltd., a language technology software company, has hired Martin Brehrov and Alex Waller as software engineers, and Chao-Hong Liu as a machine translation scientist. Iconic Translation Machines Ltd. http://iconictranslation.com
- Amplexor, a consulting services and solutions provider, has hired Marie-Laure Vinckx and Manuel Stöckl as managing directors for Germany and Switzerland. Amplexor www.amplexor.com
- Comunica, a provider of language services, has hired Nicolás Martín Fontana as vendor and sales manager. Comunica www.comunicatranslations.com
- Language I/O, a software company that provides a customizable combination of machine and human translation, has hired Fabian Lobera as chief financial officer. Language I/O LLC www.languageio.com

Resources

Annual Web Globalization Report Card

Byte Level Research, analyst of the art and science of web globalization, has released its 2020 “Web Globalization Report Card” analyzing 150 global websites across more than a dozen industry categories. It focuses on websites that have excelled in the practice of web globalization and identifies emerging trends. Byte Level Research www.bytelevel.com

Results of translator and interpreter survey, supply chains that involve freelancers

CSA Research, an independent market research firm specializing in the language service industry, has published “The State of the Linguist Supply Chain.” The report covers methodology and information sources used and presents more than 50 figures and tables with data on linguists’ offerings, background and career, clientele, technology usage, income from language services, volunteer work and the evolving future.

CSA Research has also published “Six Supply Chain Challenges.” The report is based on a survey conducted from July through September 2019, with 7,363 translators and interpreters. CSA Research https://csa-research.com

Addition to ELRA catalog

The European Language Resources Association (ELRA) has added a new pronunciation lexicon to its catalog. The SpeechTera Pronunciation Dictionary is a machine-readable pronunciation dictionary for Brazilian Portuguese and comprises 737,347 entries. Its phonetic transcription is based on 13 linguistics varieties spoken in Brazil. European Language Resources Association www.elra.info

Business

Amplexor awarded EP translation contracts

The European Parliament’s (EP) Directorate-General for Translation has awarded eight translation contracts to Amplexor — three lots as main supplier (Dutch, Italian and Latvian) and five more as an alternate supplier. The digital solutions provider also won another tender procedure for the EP as main translation partner for Spanish and Polish. Amplexor www.amplexor.com

New look for Andovar

Andovar Pte Ltd, a provider of language solutions and multimedia localization services, has updated its brand logo and colors for its four company websites. Latest offerings include AI-powered language solutions and microsites for broadcast media, eLearning and gaming content. Andovar Pte Ltd www.andovar.com

Clients and Partners

Plunet selected by oneword

Plunet GmbH, a provider of business management software for translation services and agencies, has been selected by oneword GmbH, a provider of language services, to further automate the control and documentation of its processes. Plunet GmbH www.plunet.com
oneword www.oneword.de

AppTek partners with OOONA

AppTek, a provider of speech recognition and language technology solutions, has announced a system integration partnership with OOONA, a provider of management and production tools for the content localization industry. The partnership allows OOONA to offer AppTek’s technologies within its cloud-based translation manager platform. AppTek www.apptek.com
Xillio partners with Kaleidoscope

Xillio, a provider of content migration and integration software, has partnered with Kaleidoscope, a provider of software and solutions for translation, content management and terminology. The partnership is intended to help companies exchange data between all systems that are involved in the translation and localization process.

Xillio [www.xillio.com](http://www.xillio.com)
Kaleidoscope [www.kaleidoscope.at](http://www.kaleidoscope.at)

TransPerfect Studio.NEXT and Media.NEXT Dubbing Academy

TransPerfect, a provider of global business services, has launched Studio.NEXT, a platform that uses recording-in-the-cloud technology. The company has also introduced Media.NEXT Dubbing Academy, a learning program designed to teach and develop the next generation of voiceover and dubbing talent.

TransPerfect [www.transperfect.com](http://www.transperfect.com)

Products and Services

KantanMT supports SAF-T financial standard

KantanMT, a subscription-based machine translation service, has developed a new interface for its machine translation platform that supports the SAF-T (Standard Audit File for Tax) protocol, a standard for electronic exchange of accounting data from organizations to a national tax authority or external auditors.

KantanMT [http://kantanmt.com](http://kantanmt.com)

GPI expands services

Globalization Partners International (GPI), a provider of document, software and website translation services, has expanded its eLearning Localization Practice to assist global companies in creating and publishing multilingual eLearning content.

Globalization Partners International [www.globalizationpartners.com](http://www.globalizationpartners.com)

Certifications

Literra ISO 17100 certified

Literra has been awarded ISO 17100 certification. The company specializes in life sciences, oil and gas, and military translation and interpreting projects.


Travod receives two ISO certifications

Travod, a provider of language services, has been awarded ISO 9001:2015 (quality management) and ISO 17100:2015 (translation services) certifications by TUV Austria.

Travod [www.travod.com](http://www.travod.com)
May

2020 STC Technical Communication Summit & Expo
May 17-20, 2020, Online

Content Connections 2020
May 18, 2020, Online
Acrolinx, www.acrolinx.com/cc

June

Game Global Berlin*
June 9-10, 2020, Berlin, Germany
Localization World, Ltd., https://gameglobal.events

LocWorld42 Berlin*
June 3-5, 2020, Berlin, Germany

Court Interpreter Training Institute
June 6-July 17, 2020, Online and Tuscon, Arizona USA
National Center for Interpretation, https://nci.arizona.edu/training/citi

TAUS Industry Leaders Forum*
June 9-10, 2020, Amsterdam, Netherlands
TAUS, https://bit.ly/38ALMgM

Translation, Transcreation, Transadaptation and the Science Behind Them*
June 18, 2020, San Jose, California USA
International Multilingual User Group
www.meetup.com/IMUG-Silicon-Valley/events/bplhmrybgbxb

(* = still scheduled at time of print)

July

AELFE-TAPP 2020 International Conference
July 8-10, 2020, Barcelona, Spain
Universitat Politècnica de Catalunya, https://aelfetapp.upc.edu/en

Project Underwear: Analyzing Global Consumer Expectations of Localized Content
July 16, 2020, San Jose, California USA
International Multilingual User Group
www.meetup.com/IMUG-Silicon-Valley/events/bplhmrybgbcb

UTICamp-2020
July 20-26, 2020, Dnipro, Ukraine
InTExt, https://utic.eu/en
World Savvy

Latgalian: New efforts to keep an ancient language alive

John Freivalds

John Freivalds runs an international communications firm and is the Honorary Consul for Latvia in the US State of Minnesota.

When I first became the consul for Latvia in the US State of Minnesota, right on the border with Canada, I felt confident that I knew a lot about Latvia’s regions, culture and language. Latvian is one of the EU’s 23 official languages and many LSPs offer it — there is machine translation and AI software available in Latvian.

But another language in Latvia, Latgalian, long thought to be dying, is seeing a rebirth and provides a case history on how other troubled languages and dialects can survive.

Latgale is the most southeastern region of Latvia. It is a verdant and lake-strewn place bordering Russia and Belarus. Mix those language cultures together with Latvian, stir, add a little Polish and borrow words from German, throw in a lot of “standard” Latvian and voila, you have Latgalian. Like many such languages and dialects, it’s fun to hear it spoken and try to trace the linguistic genealogy — where did that word come from?

Latgalian is classified as both its own language and a dialect of Latvian, depending who you ask. There is the cultural generalization that implies that a dialect is something less than a language. As John McWhorter wrote in *The Atlantic*: “Is a dialect, on some level, unsophisticated, as if it doesn’t have a literature because it is unsuited to extended thought and abstraction?”

When I first became the consul for Latvia in the US State of Minnesota, right on the border with Canada, I felt confident that I knew a lot about Latvia’s regions, culture and language. Latvian is one of the EU’s 23 official languages and many LSPs offer it — there is machine translation and AI software available in Latvian.

But another language in Latvia, Latgalian, long thought to be dying, is seeing a rebirth and provides a case history on how other troubled languages and dialects can survive.

Latgale is the most southeastern region of Latvia. It is a verdant and lake-strewn place bordering Russia and Belarus. Mix those language cultures together with Latvian, stir, add a little Polish and borrow words from German, throw in a lot of “standard” Latvian and voila, you have Latgalian. Like many such languages and dialects, it’s fun to hear it spoken and try to trace the linguistic genealogy — where did that word come from?

Latgalian is classified as both its own language and a dialect of Latvian, depending who you ask. There is the cultural generalization that implies that a dialect is something less than a language. As John McWhorter wrote in *The Atlantic*: “Is a dialect, on some level, unsophisticated, as if it doesn’t have a literature because it is unsuited to extended thought and abstraction?”

Do you Press 1 for Standard English and Press 2 for Appalachian Mountain English?

The Translation People blog got it right (and I paraphrase): the difference between a dialect and language is not clearly defined, is largely the result of perception and is often influenced by political factors. For example, while Dutch and Afrikaans are generally considered to be different languages, Afrikaans actually takes 90% of it vocabulary from Dutch. Conversely, Cockney and Glaswegian are considered to be dialects of English, yet speakers can struggle to understand one another.

Currently, Latgalian is accepted throughout Latvia — as a dialect. Juris Viļums, a member of parliament (Saeima), speaks Latgalian in parliament and maintains a Twitter account in Latgalian. Viļums explains in an interview with Deep Baltic that he is passionate about preserving his cultural heritage, and urges people, “you don’t have to be shy about your language.”
Latgalian has fought many battles to survive. The powers that be tried to kill it off three times and failed. First, the czars banned Latgalian from 1865-1904, and then Latgalian was removed from school curriculum by President Karlis Ulmanis in 1934. And then Latvians had the Soviet occupation (1940-1991) and cultural genocide, which included the prohibition from speaking Latgalian.

But Latgalian now has an official stamp of approval. Far from being a dying language spoken only by bent-over senior citizens, it is being picked up by a younger generation as a source of regional pride; there is even a rock band called Green Novice that sings only in Latgalian. And instead of being just mentioned in arcane academic books on the inflectional morphemes in Latgalian, you can download an app for colorful stickers to paste on your refrigerator or wherever (www.lakuga.lv). There are around 150,000 speakers of Latgalian, who now have a Latvian-Latgalian dictionary to use, and there are many other books in Latgalian.

The most useful word in Latgalian is vasals (vesels is health/good cheer in Latvian). In Latgalian it can be used for hello, goodbye and even cheers — it directly translates to “hale and hearty!”

For you Latgalian-challenged speakers, here is part of the technical description of Latgalian by noted linguist Nicole Nau, professor of Baltic languages and linguistics at Adam Mickiewicz University in Poland: “Typologically salient features of Latgalian include morphophonological harmony with an opposition of back vs. front vowels and soft (palatalized or alveolar) vs. hard consonants, a large inventory of nonfinite verb forms ... and the existence of a distinct logophoric pronoun referring to the speaker of a reported discourse.” And that’s the short description.

As Latgale resident Piters Locs said in a 2015 interview with The New York Times, “we are a separate people.” This carries not only linguistic ramifications, but geopolitical as well, and pro-Russian groups see this as a wedge to separate Latgale from Latvia. Ironically, Russia is now trying to weaponize regional languages to create divisions within countries. [M]
The speed at which software and artificial intelligence (AI) is maturing is awe-inducing. One of my favorite examples of real-world AI is Tesla Autopilot. It’s an advanced driver-assistance system that has lane centering, adaptive cruise control, self-parking, automatic lane change and even the ability to navigate autonomously on limited access freeways. But Tesla also recommends that drivers keep their hands on the wheel when using these powerful automation features.

As Tesla customers hit the pavement worldwide, the Autopilot technology is learning, and the team at Tesla is developing software and hardware to realize a future where humans do not actually have to drive because the computer will know more than we do. In fact, Tesla predicts that
Once Autopilot is proven to be 200% safer than a human, a human's input would increase risk to drivers and pedestrians. So, from an ethical perspective, one could argue that it makes sense to use Autopilot today with a human-in-the-loop approach, and once it is proven to be safer than a human, we should be ready to let the computer do the driving. My prediction is that we will look back in wonder that we let humans drive cars, much as we look back in disbelief at the use of leeches in medicine.

When it comes to translation and localization, machine translation (MT) is one of the foremost innovations that leverages AI. I'm curious to explore the ethics of MT because we are seeing such a seismic shift in how we think about AI in translation today. This has been the primary focus of interest for many practitioners and conferences in the industry for the past few years, and many companies have already started using MT, or at least begun testing it. But this is the decade where MT will become commonplace across the enterprise.

There have been major advancements and investments in MT by massive companies like Google, and growing companies like DeepL. While MT has been used quite liberally by the general public to get the gist of what a bed and breakfast offers in a foreign country, it's now becoming more commonplace by companies that might typically engage in human translation.

The most significant concern about MT is that it can be woefully inaccurate. MT is predicated on the data and content that's available, and in some cases, the content and data is inaccurate, incomplete and even gender-biased.

So there are two questions that come to mind when considering the ethics of MT. First: is the MT model being designed empathically? And second: how do people and companies use MT?

**Design thinking and MT**

Design thinking is a process for creative problem solving with the human experience at its core. It has three essential components:

1. **Empathy** — understanding the needs of those you're designing for.
2. **Ideation** — generating a lot of ideas. Brainstorming is one technique, but there are many others.
3. **Experimentation** — testing those ideas with prototyping.

The results of design thinking make answers seem obvious and natural, but the reality is that the process behind design thinking enables the user experience to feel simple. As an example, have you ever heard of PillPack? It's a medical prescription home-delivery solution that does everything from managing refills to coordinating insurance payments for their patients. They worked with Ideo to develop a brand strategy and product that sought to fully understand and embrace the customer experience. The business sold to Amazon in 2018 for a cool $1 billion.

Another company, Airbnb, was famously struggling to grow. They came to realize listings without photos weren't being booked. So, the Airbnb team started photographing listings in New York City with their customers. It wasn't long before they realized that photos offered their consumers the requisite validation for Airbnb hosts to successfully list their homes.

Design thinking has its place in MT, too. And that's where ethics comes into the picture.

When we think about MT, our perspective is that we have to input more data than one might imagine to build a baseline model. Then we use statistical data analysis to measure the accuracy of the target translations. It's at this point where empathy, ideation and experimentation come into play. This is also the point at which companies that develop MT models can establish a differentiated product.

That's because a human has to be looped in during this process to make decisions that impact the long-term development of the MT model. Statistical outliers have to be removed from the engine, reviewed by a human and edited by a human where necessary. The widely recognized issues of inaccurate, incomplete or biased data can negatively impact a person's experience with content if outliers and at-risk translations aren't addressed by a human.

Of course, new technology always introduces risk. Developers and product managers who leverage design thinking must actually put themselves in the shoes of the consumer to understand how it would feel to be wronged by the very product being developed. This is no trivial task! Leaning on the human-in-the-loop methodology lends itself to building a better MT model and a competitive advantage in a crowded field.

**How companies and people use MT**

For decades, consumers have relied on MT engines to achieve a basic understanding for a worldwide product, service or information. Sometimes it works great. Other times the translations are questionable, or even downright dangerous. I'm sure you wouldn't bet your life, or someone else's, on the accuracy of MT if you didn't understand the language.

This analogy can be used as a guiding principle for people and companies considering MT for their content today. If the translation is
mission-critical or can put a life at risk, it’s worth investing in human translation.

That’s why Translators without Borders enlists human translators and interpreters to realize humanitarian communications worldwide. They’ve translated over 83 million words. Some quick math here: 83 million words times six characters (average number of characters per word) is about 500 million characters. That’s about $10,000 in MT — a lot less than the actual value of the professional translations donated by Translators without Borders’ 30,000 human translators!

But there’s a good reason for this: people’s lives are on the line. Some are seeking citizenship, others are facing life-or-death humanitarian crises. It would not be ethical to depend on MT when the probability of an error introduced is significantly higher than if a human translator is assigned to the case. It’s not worth the toll on human life.

Another example: when the deadly Covid-19 (Coronavirus) spread in Wuhan, China, foreign governments donated medical equipment and documented procedures to help contain the epidemic. As reported by Global Times, a Chinese publication that is translated into English, volunteer translators chipped in to translating a word count of 272,600 characters (this is about 409,000 words in the English language). It would have been faster and less expensive to machine translate the documents and product descriptions, but it would not have yielded the intended results. Lives were on the line, and using MT for this work would have been unethical.

MT is a perfectly suitable solution for a lot of different content types, and it will become better for all content types in time. Companies today frequently use MT for lower-value content: certain help content, internal communications, low-priority texts that are infrequently read by consumers. If someone’s life isn’t at risk, the decision to use MT is largely a business calculation. How much risk are you willing to take for the potential reward?

Finding balance

How we think about technology, artificial intelligence and machine translation in our industry is a shared responsibility, just as Tesla and its owners consider its Autopilot a substantial innovation and achievement, but that still requires a human-in-the-loop approach.

New technology always brings with it unknown ethical challenges. We can be mindful of these challenges in two ways. We can engineer these products with a design-thinking approach: consider the end-user, and work with a human translator to actively improve the language learning model. Additionally, we can consider how and when MT should not be used: if the end-user’s life depends on the content, do not depend on MT. [M]
Keynote Speaker:
Larry Hochman

Larry Hochman is one of the most popular keynote speakers in the world, having given more than 500 speeches in 72 countries. Larry is a former European Business Speaker of the Year, as well as one of the world’s most influential advisors on building organizations that deliver an exceptional customer experience.

June 3-5, 2020
InterContinental Berlin, Germany

OR ON YOUR SCREEN

The world’s #1 Localization Conference & Exhibition Series

Engage Global Users

Discover conference and preconference programs, networking events and more at locworld.com/locworld42

We are monitoring the COVID-19 situation. Please stay healthy and check our website for the latest details.
The dawn of a new linguistic epoch

Time is a funny thing. It’s hard to believe five years have already elapsed since the promising early results of applying deep learning and artificial neural networks to machine translation (MT). Yet it’s equally hard to believe it’s only been that long — that a radically new method of approaching one of computer science’s oldest challenges, automated translation, could so quickly and completely usurp statistical machine translation.

Since our company started doing post-editing at scale in 2001, we now have hundreds of millions of words that allow us to track trends in MT quality and the amount of changes that professional translators have to make to MT output to create quality translations. The data confirms what readers likely already know: carefully trained neural machine translation (NMT) produces draft translations that require substantially less editing.

That’s a very simple and straightforward statement, but one with potentially massive implications for our industry — and that’s just the first concrete and most obvious example of applying artificial intelligence (AI) to language service providers’ offerings.

Let’s pause there for a brief word on AI, the buzziest of all buzzwords. Whenever AI is mentioned, it’s important to establish what the term actually means. AI systems fall into three categories:

- Systems designed by people and trained with people-created data that perform a task assumed to require intelligence.
- Systems designed by people that can iteratively improve their own task performance without human intervention.
- Artificial general intelligence systems that were initially designed by people but now have both the ability to independently improve performance and autonomous decision-making abilities — such as changing tasks or self-improvement.

NMT, like most AI products in use today, falls squarely into the first category. This means in terms of analyzing mountains of data and applying those learnings to a task, AI is closer to children’s book character Mary Anne, a steam shovel that “could dig as much in a day as a hundred men could dig in a week” than to the humanity-attacking machines from bad science fiction.

That doesn’t mean, however, that NMT and current AI aren’t tremendous achievements. They are, and they bring with them challenges and opportunities we’re just beginning to come to terms with. But responding to change is nothing new in an industry as old as ours. Language service providers have witnessed and adapted to every major technological advance since translators moved from etching characters into clay tablets to writing them with ink on parchment. What is new is the rapidity of change that we’re currently witnessing. Think of the changes that a translator approaching retirement has seen over the course of that career, witnessing the digitization of nearly every aspect of the job. She might have started translating on a typewriter, then moved into a word processing program, then to specifically designed...
translation environments, having to go through the massive rethinking and payment changes that came with translation memory. Her stack of dictionaries may have been replaced by computer aids like spellcheck or other online resources. By the time our industry figures something out, it can seem like the technological landscape has already moved on. And today we see MT producing output that is quite often shockingly good.

Change is inevitable, but how we confront it is within each individual’s power. When I try to forecast what’s going to happen in our industry regarding NMT and AI, I find it instructive to look to history for similar events to see if there are lessons to be learned. Three very different historical scenarios come to mind, each illustrating ways in which technological change has had radical implications for those involved.

1. Typesetting

Upon its establishment in 1852, the National Typographical Union — later the International Typographical Union — became the first trade union in the United States. It was founded to protect and promote the tens of thousands of workers who manually set type for newspapers, magazines and book publishers. In its heyday, the union had more than 120,000 highly skilled members — people who had the education, skill and dexterity necessary to read a manuscript and render it in reverse with stunning speed and accuracy. But mechanization and automation radically reduced the need. The union disbanded in 1986, just about the time anyone with a laser printer could produce camera-ready copy. Today in the age of ebooks and online newspapers, the occupation no longer exists outside traditionalists who practice typesetting as an art.

2. Acting

When the first feature-length movies were produced more than 100 years ago, before filmmakers had the technological ability to add synchronized soundtracks, any actor who wanted to be successful had to adopt extreme mimicry to convey emotion. A scant twenty years later, for a generation of silent-era actors changed when 1927 saw the launch of talkies — movies that had sound. Speech coaches were suddenly in great demand as enunciation — a skill movie stars had never needed — was suddenly critical. Even stage actors of the day, who had never forsaken their oratory skills, had to modify their diction, which had been completely unnecessary on a sound stage. Technology-driven market changes demanded they expand their skillset by working on enunciation and the expressiveness of their voices. Among those silent-film stars who were able to adapt to voice acting are some of the biggest names of Hollywood: Greta Garbo, Joan Crawford and Carole Lombard. But many others, like Mary Pickford, Douglas Fairbanks and Charlie Chaplin, never became comfortable with voice acting and their careers suffered for it.

While it’s easy to focus on actors who were left behind, adding dialog and sound created far more jobs than it cost — jobs that didn’t exist before such as sound technician, boom operator, sound effects specialist and sound mixers. And of course the screenwriter’s job changed radically.

To give you an idea of how quickly talking pictures became the norm, consider this: at the first Academy Awards in 1929, the films nominated for Outstanding Picture were all silent. But the next year, every single nominated film had sound.

3. Field hockey

A German friend of mine has been playing field hockey competitively since the late 1970s. Over the course of his involvement, international field hockey has gone from being played on natural grass to artificial turf. On artificial turf, a field hockey ball rolls much faster and further than on natural grass. It can be struck more cleanly and is subject to fewer unexpected hops and direction changes. The result is that today’s game is much faster and more precise with an emphasis on long passes and terrifyingly hard shots. Winning on artificial turf requires modified skills — particularly the ability to trap speedy passes, new offensive and defensive strategies, greater levels of fitness and even changes in equipment with modifications from shoes and sticks to considerably more padding for the goalkeeper.

So yes, the translation and localization game is changing quickly and it is up to each of us — whether we work independently or as part of a corporation — to ensure that we have the proper skills, equipment and strategies to play competitively and successfully.

Here are some of the changes to our game that I consider likely, presuming no further dramatic leap in the sophistication of machine translation systems:

| Translation memories will be used for 100% matches and for MT training material, but their importance for partial matches will diminish. |
| Translation of consistently authored informational content will become a review process of substantially accurate and fluent machine-generated translations by professional translators with subject area expertise. |
| Unsupervised MT — MT output that has not undergone human review — will be used much more... |
broadly and confidently as a completely adequate solution for lower-risk content.

Increasing volumes of MT output that has been carefully post-edited will lead to continuing incremental improvements in MT quality.

Significant improvements in efficiency will lead to a rethinking of pricing away from per-word and toward pricing based on business value and the level of necessary human interaction.

Big data analytics will allow new ways to analyze how well a translation performs for the customer, meaning effective resolution of a business need will be the measure of translation quality, which in turn will allow a more meaningful tracking of the ROI of a translation.

Rethinking of pricing and clearer ROI models will lead to substantial increases in translation volume.

Some of these changes are already taking place, while others will require a considerable amount of time. In an industry as fragmented as ours, we can also be sure that these changes will not happen uniformly. As cyberpunk novelist (and hero of mine) William Gibson said in an interview in The Economist in 2001, “The future is already here, it’s just not evenly distributed.”

The effects of technological advances on the language services industry don’t start and end with NMT. The next epoch is already upon us. Our industry tells a much broader story and the continuation of that story represents massive opportunity. For millennia, we have been translating texts and interpreting speech so that they can be understood by people who speak different languages. For decades, we have been localizing products so that they are accessible to users around the globe. And now, in the age of AI, there is a vast and growing array of smart products that need linguistic intelligence, so that they can understand our speech and so that our wants and desires are accessible to them. Not unlike the introduction of speech to motion pictures, adding linguistic functionality to smart products has created new opportunities in the creation, curation and testing of the linguistic data sets from which these systems learn. This is a whole series of language services that didn’t exist a few short years ago because the demand for them didn’t yet exist.

Even as we work to create ever more effective methods of translating texts, interpreting speech and localizing products, the next quantum leap is here. We are no longer just localizing products so that we can understand their interface but so they can understand ours. That’s a big part of what AI is and it’s a growing part of what the language services industry is.

So while our industry seems to be in a head-spinning state of change, it’s critical to remember that we have been reacting to change for decades and decades. What hasn’t changed — and what should never change — is our focus on fulfilling the multilingual needs of our customers. Those needs will continue to change as the world around them changes. [M]
For twenty-nine years the Internationalization & Unicode® Conference (IUC) has been the preeminent event highlighting the latest innovations and best practices of global and multilingual software providers. Please join us for our 44th conference! This year’s event is being held October 14-16, 2020 in Santa Clara, California. Recent conferences have provided specific advice on designing software for European countries, Latin America, China, India, Japan, Korea, the Middle East, and emerging markets.

The Internationalization and Unicode Conference® (IUC) is the annual conference of the Unicode Consortium where experts and industry leaders gather to map the future of internationalization, ignite new ideas and present the latest in technologies and best practices for creation, management, and testing of global web, and multilingual software solutions.

Reasons to Attend Include:

- Tutorials and sessions for beginners, to train you and your staff on basic practices and implementation techniques for creating international software
- Learn recommended solutions to difficult problems or sophisticated requirements from industry leaders and experts in attendance
- Find help from tool and product vendors to get you to market quickly and cost-effectively

MEDIA SPONSOR:

For up-to-date information or to register:
VISIT: www.unicodeconference.org/ml
E-MAIL: info@unicodeconference.org

Unicode and the Unicode Logo are trademarks of Unicode, Inc. All other marks are the property of their respective owners.

www.unicodeconference.org/ml
Since its foundation as an academic discipline in 1955, the AI research field has been divided into different camps: symbolic AI and machine learning. While symbolic AI used to dominate in the first decades, machine learning has been very trendy lately, so let's try to understand each of these approaches and their main differences when applied to natural language processing (NLP).

Machine learning is an application of AI where statistical models perform specific tasks without using explicit instructions, relying instead on patterns and inference. Machine learning algorithms build mathematical models based on training data in order to make predictions.

Machine learning technology uses algorithms to teach the computer how to solve problems, and how to gain insights from solving those problems. That’s how the computer learns automatically, without human intervention or assistance: by observing and looking for patterns in data and using feedback loops to monitor and improve its predictions. While humans would be overwhelmed with masses of data, machine learning thrives and is able to evolve its understanding in order to make better decisions in the future, based on the examples that were provided to it.

**Machine learning applied to NLP**

Machine learning can be applied to lots of disciplines, and one of those is NLP, which is used in AI-powered conversational chatbots.

Here’s how machine learning works in this specific case: the person who oversees the bot, usually called a botmaster, feeds the engine with as much relevant data as possible. The bot then gets asked questions by its users and it automatically decides which answer to push for every intent it’s queried for. An intent, in this context, is a kind of baseline query. You can type “show me today’s news” or “what’s the news today?” and the bot should recognize that the intent is the same.

The botmaster then needs to review those responses and has to manually tell the engine which answers were correct and which ones were not. That is how the machine learns how to serve the correct answer.

As you can easily imagine, this is a very time-consuming job, as there are many ways of asking or formulating the same question. And if you take into account that a knowledge base usually holds on average 300 intents, you now see how repetitive maintaining a knowledge base can be when using machine learning.

Don’t get me wrong, machine learning is an amazing tool that enables us to unlock great potential and AI disciplines such as image recognition or voice recognition, but when it comes to NLP, I’m firmly convinced that machine learning is not the best technology to be used.

**Symbolic AI**

Symbolic AI, also known as good old-fashioned AI (GOFAI), uses human-readable symbols that represent real-world entities or concepts as well as logic (the
mathematically provable logical methods) in order to create rules for the concrete manipulation of those symbols, leading to a rule-based system.

In a nutshell, symbolic AI involves the explicit embedding of human knowledge and behavior rules into computer programs.

One of the many uses of symbolic AI is with NLP for conversational chatbots. With this approach, also called “deterministic,” the idea is to teach the machine how to understand languages in the same way we humans have learned how to read and how to write. In order to do so, we went to school and we learned how to structure language through rules, grammar, conjugation and vocabulary. Computational linguists do exactly the same: they use rules, lexicons and semantics in order to teach the bot’s engine how to understand a language.

Using symbolic AI, everything is visible, understandable and explainable, leading to what is called a “transparent box,” as opposed to the “black box” created by machine learning.

As a consequence, the botmaster’s job is completely different when using symbolic AI technology than with machine learning-based technology, as the botmaster focuses on writing new content for the knowledge base rather than utterances of existing content. The botmaster also has full transparency on how to fine-tune the engine when it doesn’t work properly, as it’s possible to understand why a specific decision has been made and what tools are needed to fix it.

To summarize, one of the main differences between machine learning and traditional symbolic reasoning is how the learning happens. In machine learning, the algorithm learns rules as it establishes correlations between inputs and outputs.

In symbolic reasoning, the rules are created through human intervention and then hard-coded into a static program.

If machine learning can appear as a revolutionary approach at first, its lack of transparency and a large amount of data that is required in order for the system to learn are its two main flaws. Companies now realize how important it is to have a transparent AI, not only for ethical reasons but also for operational ones, and the deterministic (or symbolic) approach is now becoming popular again. [M]

Connect the dots to open up a universe of opportunities

The language industry is changing, driven by the changes brought about by the content economy. With Smartcat, you have the tools to explore the new horizons. Are you ready to take it to the next level?
AI in the translation business: Beyond MT

Dmitry Ulanov has been involved in localization since 1998 and risen through the ranks from a project manager to CTO. Now he is focused on automation of operations and also supervising the R&D department. He holds an MS degree and is an accredited internal auditor for ISO 9001 QMS.
There are many areas in the translation business that may benefit from implementation of AI and particularly from machine learning (ML). Since AI is not completely predictable, implementation of any kind of AI also involves new risks. The introduction of AI may produce an unexpected impact or a positive breakthrough.

In linguistics, we have been seeing a continuous evolution of machine translation (MT), the most evident example of ML. We already know what to expect from it, what the benefits are, and how to address the risks. How we manage it also presents a great potential for applying ML to make improvements elsewhere. This includes automated decision-making based on acquired data to either provide human workers with pointers, or to eliminate certain human management in order to resolve bottlenecks in continuous development or continuous translation. We could also use ML for making predictions and bringing to light potential risks (of missing a deadline, losing a client) and maybe to propose corrective measures.

There are different areas to which an ML technique could be applied, and different learning approaches — supervised or unsupervised learning, reinforcement learning, transfer learning and so on. Searching for the best combinations is an interesting journey.

First steps: ML-powered vendor suggestion

A couple of years ago, we conducted a survey among our project managers to find out in which areas the use of ML could simplify and speed up their work. It turned out that the project manager spends quite a lot of time searching for the right vendor to do a particular job, and this is far from the project manager’s favorite activity (Figure 1). This was why we began experimenting with ML to select the best vendor for a project.

To begin with, we wanted to find out how realistic it was to train a neural network to make the same choices as human project managers. We took the initial data and assigned vendors from several tens of thousands of already completed jobs and used them to train a model. Although a variety of sophisticated neural network architectures are available, we chose a simple regression classification model, and it performed fairly well. Validation showed that the model made the right choice in more than 99% of cases. However, when trying to use this model’s output in production, we found that often, selected vendors were in fact not suitable. This was due _inter alia_ to the fact that in the past, work was often given to a vendor that would be far from the best one for the job from today’s perspective. It was the principle of “garbage in/garbage out” in action.
<table>
<thead>
<tr>
<th>Manual activity</th>
<th>How to partially automate with ML</th>
<th>How to eliminate the manual activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Looking for resources</td>
<td>Suggest resources (vendors, TM, glossary, etc.) from similar projects</td>
<td>Choose the best vendor for each job considering language pair, domain, qualification, experience, cost, availability, productivity, reviewer feedback and other parameters.</td>
</tr>
<tr>
<td>Project set up and planning</td>
<td>Propose project workflow according to SLA and similar past projects; estimate task ETAs and critical path</td>
<td>-</td>
</tr>
<tr>
<td>Manual preparation of vendor instructions</td>
<td>Extract key requirements and draft structured vendor instruction based on them</td>
<td>-</td>
</tr>
<tr>
<td>Deadline monitoring</td>
<td>Estimate probability of not meeting deadline in real time based on historical data + project manager alerts</td>
<td>A few control layers could provide reliable deadline control to eliminate manual efforts altogether</td>
</tr>
<tr>
<td>Correct data in the system</td>
<td>Background analysis of data consistency based on common patterns + alert responsible employee about possible wrong data</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 2: Charting out improvements based on machine learning.

Reevaluation of project management

Recently, we conducted another survey to understand what activities project managers spend the majority of their time on. In addition, it was interesting to learn in which other areas it makes sense to experiment with machine learning. This survey showed that the search for the best vendor for a project is still the number one task. It also became clear what other areas needed work. Figure 2 lists activities of project managers we are going to improve and ideas about how to partially automate this type of activity, or completely pass it to the machine.

Vendor suggestion 2.0

As mentioned above, the ML-based vendor suggestion worked, but often gave results which were wrong (or at least seemed to be wrong to some of the project managers). To resolve this, we decided to:

1. Consider who is managing this project (and therefore who receives a suggestion). This allows customization of ML output on a personal/departmental basis, and thus, to continue reusing collected data across different project managers and eliminate “wrong” suggestions at the same time).

2. Reduce the weighting of records in the training/validation data set as they become obsolete. Thus, more recent data take precedence over data that is five years old, and the data older than five years is not considered at all.

The results of using this improved architecture have not yet been announced, but we already have a list of adjustments for the next iteration. Instead of a complete regular retraining of the model on the data set over the past five years, we will try to switch to reinforcement learning on the data that will come after the initial training of the network.

Apart from that, we can just compare a new project with completed past projects considering such characteristics as client/account, language pair(s), domain, service level agreement and a linguistic snapshot of text to be translated (hello, natural language processing). And we simply reuse resources from matching past projects where appropriate.

Correct data in the system

While processing data contained in the records of completed jobs, we found a number of inconsistencies. Analyzing these, we realized that with a little extra effort it is possible to identify incorrect data with a high degree of certainty — the existence of which also poisons the lives of project managers. The appearance of incorrect data could be due to a wrong user action, an error in the program code or just a trivial typo. No matter the cause, at least we could easily identify it.

So, the way to detect erroneous data could be either an extra analysis during a scheduled data processing task or a separate scheduled data analysis task. Both are supposed to identify cases that do not match a common pattern (via regression model). Automatic correction of data that seem incorrect would be too risky. Therefore, it is enough to report a potential error to the responsible employee who can verify and fix it if needed.
Project setup and planning

Often at the start phase of a project, such as quotation or project acknowledgement, it is necessary to determine a realistic project deadline. To get an answer to this question, we need to evaluate the deadline for each stage in the project workflow, calculate the critical path and add some time buffer for contingency. The estimated deadline for each specific task can be found by a neural network that was previously trained on data from completed tasks.

There is another option that would work for a quick preliminary estimation of the deadline: training the network on the data of previously completed projects, taking into account the volume, language pairs, domain and other characteristics of the project. The estimate given by such a network will be approximate, but such an algorithm is much simpler to implement.

Deadline monitoring

For project deadline monitoring, we can use the same approach as for project planning. However, when the project has already been launched and work is in full swing, we have much more data about it. For example, it is known which tasks have already been completed and which specific vendors are assigned to the remaining tasks. Accordingly, in this case, it is enough for us to evaluate the realistic completion time for the remaining tasks, calculate the critical path, add it to the current time to estimate the most probable project completion time and compare it with the agreed deadline. If the estimated completion time is earlier than the agreed deadline, we are happy — and there is every chance we will meet the deadline. And if it’s later, then by applying the difference to a suitable normal distribution histogram, you can roughly estimate the probability of still meeting the deadline (which in this case will definitely be lower than 100%).

Manual preparation of vendor instructions

Although preparation of clear, consistent and up-to-date instruction for vendors is an important part of the project management routine, it often ends up in the copy-and-paste of client instructions, which is crude and unrewarding. While maintaining a set of up-to-date instructions for all types of vendors (like engineer, translator, editor, proofreader and DTP specialist) involved in the workflow for a certain account is a good practice, this could be excessive or just not realistic in some cases.

Here to help us are natural language processing techniques such as information extraction and named entity recognition. Running them on a project description received from a client in nonstructured format, we could get all those project properties (such as volume, language pair, account, project name, deadline and many more) extracted as separate values. Then we just fill in a structured instruction template with them to get a draft of the vendor instruction. And if you do not have a corresponding project created in your project management system yet, it may be a good idea to create a project based on the extracted project properties.

Build or buy

As technology moves forward, and the world becomes more and more digitalized, software development becomes more and more important to linguistic service providers. However, complex, relatively new and quickly evolving areas like ML could still be too difficult to adopt.

To deal with this, a company may consider either building its own ML microservices with Python/R or use excellent cognitive services provided by such IT giants like Google, Amazon and Microsoft, or integrators like Intento (Figure 3).

Although the results of adopting ML could be amazing, they are accompanied with risks. So it makes sense to implement ML solutions in an operational environment, but to keep a human eye on what they do until you can trust the automation and properly evaluate the possible risks. [M]

<table>
<thead>
<tr>
<th>Machine learning services provider</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local proprietary development</td>
<td>• Better understanding of what happens under the ML hood • Absolute confidentiality • Better availability</td>
<td>• ML development team may be too expensive for a small/mid-sized company • Development may take longer • Product quality may be lower</td>
</tr>
<tr>
<td>Professional cloud service</td>
<td>• Quick start • Proven quality solutions • Higher peak performance</td>
<td>• Disclosing of data to a third party • More expensive in the long run</td>
</tr>
</tbody>
</table>

Figure 3: Weighing the pros and cons of different ML approaches.
How AI is transforming the content supply chain... and your job

Strategic account director at SDL since 2014, Jessica Roland works with selected enterprise customers to help them reach global audiences and enhance customer experience, increasingly via AI. Prior roles include leading enterprise software globalization teams and international product management.
There are many ways we are seeing AI used today in our customer and partner ecosphere. Localization professionals will be familiar with some of the basic AI building blocks and current applications, but there may be new ideas for AI paths you may want to pursue in the future, whether for your current job or a future one. After all, “AI is going to have as big an impact on our society as electricity,” according to Risto Siilasmaa, chairman of the board at Nokia.

Figure 1: Common AI-related terms.

In Figure 1, we show some of the most commonly used AI-related terms and how they fit together. Computer science is depicted as the general study of computers and computing concepts. Although the exact definition of AI is a matter of debate, for this discussion let’s define AI as systems that can perform complex tasks autonomously and adaptively; that is, without constant guidance by a user and with the ability to improve performance by learning from experience. Machine learning is a process by which systems improve their performance in a given task with more and more experience or data. Deep learning is a class of machine learning algorithms that uses multiple layers to calculate high-level outputs from raw inputs. With the increasing power of modern computers, more and more layers are added for greater complexity.

Neural machine translation (NMT) is an AI application that uses machine learning of the deep learning type to perform automatic translation tasks. NMT is the example of AI that most localization professionals are aware of today. Its ROI is clear, its adoption is growing rapidly, and it is a safe and helpful enterprise tool as long as it is deployed securely.

Now that we are clear on the vocabulary, let’s see how AI enables a machine-first approach that can have real human-first impact. Today, AI is already breaking down language barriers that stand in the way of communication and collaboration, eCommerce, customer support and ediscovery. How? Let’s look at some real-life AI stories.

One of the most well-known US designers and producers of semiconductors and integrated circuits had a problem. With an increasingly global workforce, key research and development (R&D) teams had been located overseas, speaking different languages but still needing to communicate and collaborate closely with colleagues back in the US. Some of the overseas workers, while comfortable collaborating with their on-site colleagues, had trouble in written communication exchanges with the US teams, and vice versa. It became clear that translation would be helpful in ensuring both sides fully understood each other. The company already had a translation management system in place and a solid translation supply chain. But of course, the speed of putting text through a traditional translation workflow was not conducive to real-time collaboration, and the volume of communications made a traditional approach cost prohibitive. Real-time machine translation was a natural solution.

This was successful from the start, but even more so today, now that AI-based NMT is available. Using deep layers of nodes between input and output layers, NMT achieves a greater amount of accuracy and fluidity than previous generations of translation systems. Deployed within a system that keeps all the translation queries and results within the company’s firewall, NMT has enabled this manufacturer’s research teams to communicate accurately and instantaneously, increasing their R&D productivity while protecting their intellectual property.

NMT is making its way into many other domains. The eCommerce space is fiercely competitive, and time is of supreme value here. One of the largest online travel companies in the world determined that a key part of their competitive advantage was in instantaneously adapting incoming and updated hotel property descriptions into other languages, at low cost and very high volume, using NMT. Same for a household name in online retail, which, in addition to ensuring product descriptions are updated in real time, also uses NMT to enable their very large global workforce to react quickly to changing corporate guidelines. They find the best way to do that is to send instructions in the employees’ own language.

The fluidity of NMT is especially helpful when humans are trying to have a live conversation. Many companies are exploring AI chatbots as a way to simulate a human conversation in support of conducting day-to-day business. One of the most well-known US banks is deploying a sophisticated multilingual chatbot system to ensure that non-English speaking customers fully understand information they are
receiving. This is very important to them not only from a customer satisfaction point of view, but also for legal protection, to avoid accidentally landing in a fraudulent situation.

Also in the legal domain, NMT is making a huge difference in ease of ediscovery. Many government and corporate organizations are faced with an insurmountable volume of incoming multilingual data that would be too slow and cost-prohibitive to translate with traditional methods. For government agencies, secure NMT (not indexed by internet search engines) is critical to international diplomacy and security.

But as we know, NMT is just one example of AI applications that are currently in use today, both commercially and experimentally. Figure 2 shows other applications of the technology.

Credit card fraud detection, self-driving vehicles, personalized web content and facial recognition are all in the news today. They catch our attention as human beings anticipating, often apprehensively, a highly automated, unknown future. But for localization professionals, in our day-to-day work, are there other applications of AI beyond NMT that are actually real and emerging in practical use today?
Before we get to that, let’s leap forward into the localization future that we all can imagine, but that is not quite here yet on a practical level. In the same way that AI has drastically changed the world for translators, the collective localization-sphere expects to see AI transform the world of the localization project manager. There are many problems in the localization project world that would be aided by faster and more accurate, less human-error prone decision making or status monitoring. Autonomous and adaptive methods of assigning language service providers and assigning individual translators, reviewers or other project resources are all easily imaginable. So is predicting delivery dates and quality, and predicting issues in those areas for specific jobs, at a high volume with low human touch. Adaptive and autonomous job routing to maximize quality and on-time delivery, while minimizing cost, is a clear and desirable application from the enterprise perspective. Localization project managers’ jobs will be enhanced, as they gain the ability to handle job volumes that today would be overwhelming. Even one level up from the actual project management, at the data aggregation and reporting level, one can imagine this becoming highly automated and adaptive, and low-human touch (Figure 3). With human efforts augmented by machine automation, project managers will have more time for localization strategy definition, project exception management and tools strategy.

Whether it’s translation output, translator performance data or project management data, one thing is certain: your data holds the secret to your success. Companies need to think carefully about deliberate sharing of this data and be extremely wary of inadvertent sharing via nonsecure tools.

Coming back to actual emerging applications, an insight has already started to alter the work of many localization professionals: that translation is just a step in the overall content process of creating, transforming and distributing content. In order to deal with the exponentially increasing demands for producing more content — continuously, faster and more personalized — there is no place for friction in the content supply chain, including during global transformation (Figure 4).

As part of surviving in an accelerated content supply chain, localization professionals need to be aware that AI is also presenting itself in applications that today are able to streamline content creation, management and distribution. For example, the very same linguistic building blocks that are used to create NMT are also helping transform the professional lives of content creators upstream.

As an example, a well-known worldwide entertainment company and long-time customer for both web content management and translation management was traditionally known as a master communicator always on top of their global marketing game. Yet their digital marketing team was starting to feel crushed by the weight
of continual demands to create ever-increasing variations of content for social media postings, blogs, website blurbs and other publicity channels. The required content was no longer creative — it was just a reformulation of existing content. Human workers who had previously enjoyed the work of content transformation were simply becoming numbed by the sheer volume of transformation required.

It turns out this is a perfect application for AI, given the existence of the above building blocks. An original longer piece of content can now be analyzed, summarized and transformed into “snackable content” that is consumed in shorter formats such as tweets or LinkedIn postings, but is all part of the same messaging.

AI presents the content creator with the available AI transformation result and provides tools for the human to adjust the AI output, automatically calling out aspects the human might want to highlight. AI permits a much faster reaction to personalization requirements. Effectively supporting a web content personalization process has resulted in marketers achieving, on average, a 19% uplift in sales. This is just one example of how linguistic AI can influence website functionality. Search and predictive text are other well-known areas.

For localizers who support marketing or documentation teams, the accelerated creation of short-form content variations will make implementation of continuous localization and use of NMT even more important.

We’ve seen that linguistic AI can serve both localization and content management purposes, with human-plus-machine being able to accomplish more and faster, together. There are several notable examples of localization leaders who have expanded into content management, as the two fields are have a common touchpoint in words. Linguistic AI is extremely effective in both domains — already used in current applications such as NMT and emerging in content creation assistance. Being familiar with current and emerging AI applications; thinking ahead to what types of problems lend themselves to autonomous and adaptive solutions; and protecting company data that is the lifeblood of AI can help localization professionals expand beyond their current horizons and increase their contribution to their company’s success in the exciting and challenging times to come. [M]
Multi-engine MT management for global companies

Marco Zappatore, telecommunication engineer, works as adjunct professor of CAT Laboratory at the University of Salento (Italy). His research activities focus on data management, wireless propagation, computer-assisted and machine translation, as well as video game localization.

Filip Šanca is a graduate of Charles University in Prague. Having successfully established the academic partner network at Memsource, he is now coordinating the Memsource Certification Program and leading the marketing department.
Machine translation (MT) is a constantly evolving field, where research contributions from natural language processing, computational linguistics, artificial intelligence and computer science merge together with the aim of improving the quality and reliability of automatically translated texts. Several different MT approaches have showcased their advantages and limitations so far, thus motivating the adoption of multi-engine MT (MEMT), where multiple MT systems are queried at the same time in order to provide the best translation alternative.

However, significant costs, scarce automation and the risk of nonoptimized configurations make this promising research area less of a viable solution for global companies. There is a need for MEMT management systems capable of autonomously suggesting the optimal MT engine for the given language pair and domain to a translation company.

The research in the MT sector is experiencing continuous growth and diversification. During the last decades, several MT approaches have been examined, starting from the widely-known rule-based MT (RBMT). RBMT exploited a sequence of analysis, transfer and synthesis components to obtain a target language representation of the original text from a set of grammar rules. However, RBMT required adapting its components to each language, thus achieving limited effectiveness. Statistical MT (SMT) then gained relevance. SMT systems embed a translation model that is trained statistically on parallel bilingual text corpora, meaning it infers statistical translation models from human translations already available. It follows a language model that provides the knowledge of correct target language phraseological structure. Although more efficient than RBMT, the SMT approach has some limitations, as it requires a high similarity between the texts to be translated and corpus texts, as well as huge training corpora that must be correctly aligned and properly balanced. This makes SMT less suitable for less-resourced languages.

Further improvement has been introduced with neural MT (NMT), where multiple (hidden) layers of interconnected processing nodes (a neural network) are fed and activated with input language patterns (training material) so that it is possible to predict the likelihood of the target sentence structure. NMT requires considerably less memory than SMT, but its effectiveness is directly proportional to the availability of training material. It thus demands additional efforts such as a pivot language when under-resourced language pairs have to be dealt with. The inherent complexity and lexical/syntactic ambiguity of natural language, however, pose several challenges to MT approaches, such as homonymy, polysemy, homography, sentence structure, anaphora, idiomatic expressions and so on.

In order to tackle these issues, hybrid MT (HMT) techniques have been investigated aiming to improve translation quality and accuracy. MEMT represents a promising HMT technique where several MT systems provide their own translation for a given source text (typically via online APIs) and then their output is properly combined. Several hybridization methodologies have been explored so far. These can be applied once each MT system has produced its translation ranked selection. They might use BLEU scores; a one-to-one combination of translation fragments using confusion networks; or many-to-many combinations of translation fragments using lattice-based networks.

Combining the output of different MT systems at the sentence, segment or fragment level brings a significant qualitative improvement, but cannot be considered as the long-awaited silver bullet for translation companies. Indeed, MEMT-based approaches that exploit multiple MT APIs require a company to pay for more than one MT usage plan at a time but, actually, they use just a small percentage of the overall text amount the company paid for. When we count in the ROI aspect, then, the MEMT approach mostly makes sense on the job or file level if you can, on one hand, evaluate more engines at once but, at the same time, pay just for the one translation you use in the end. There are existing tools that already enable companies to implement this approach, while automating the MT selection process on the basis of MT quality scores and thus minimizing the manual effort.

The global MT market

Technology advancements in AI and deep learning are improving MT quality significantly by reducing post-editing efforts. Higher productivity rates and lower costs, if compared to human translation, are making MT an appealing choice for a continuously growing number of language combinations. Suitable content domains are extremely diversified and span from healthcare to marketing, from IT to automotive, from defense to education.

Such a favorable technological landscape finds one of its key drivers in business globalization. The more companies expand their customer base and their market segment abroad, the higher the request for fast and reliable content localization becomes. The global MT market is expected to expand at a compound annual growth rate
of 19% and with a year-over-year growth rate of 15.3% between 2020 and 2024. The estimated incremental growth is from $400 million, the MT market size in 2016, to $1.5 billion, the MT market value projection for 2024. This is especially true in sectors like video content translation.

Current forecasts, however, still depict a very fragmented market for the period 2020-2024. Technavio Research predicts in their report “Machine Translation Market by Application and Geography - Forecast and Analysis 2020-2024” that big players (such as Microsoft, Google, Amazon, IBM, SDL, SYSTRAN, DeepL and so on) will widen their MT service offerings and cope with MT quality improvement, but will keep coexisting with smaller vertical realities and open-source MT services.

Selecting the MT solution capable of providing the best translation quality depending on the required language combination and the targeted domain is expected to become an essential requirement for global companies. Recent dedicated reports have assessed stock MT services on a variety of configurations in order to highlight where a certain MT engine performs the best, but a comparison of their usage patterns by global companies has yet to be addressed.

Picking the right MT engine: A real test case

In order to provide useful insights on how companies actually exploit MT services, and how their usage patterns can be improved, let’s compare MT usage data in three selected NMT engines used at Memsource. Real engine names are not shown and only a fictional label is assigned to each of them (such as NMT Engine A) as the main purpose is to show how such engines are exploited and how their usage can be improved.

The scatterplot depicted in Figure 1 shows NMT engine performances by language combination and by

![Figure 1: NMT engine performance by usage percentage and by language combination.](image)
usage. Five analysis variables are considered in Figure 1. On the horizontal axis, the percentage of usage is reported (that is, the ratio between the number of MT-translated segments and the overall number of segments in a given language combination) while on the vertical axis the average MTPE score is shown. Data points are:

1. shaped in such a way that each NMT engine is represented with its initial (so A stands for NMT Engine A);
2. colorized according to the examined language combinations (so EN>CS is in green);
3. sized depending on the overall number of segments for each language combination (so a bigger initial corresponds to a higher number of available segments).

For instance, the small orange A in the upper left reveals that the first NMT engine has an average 0.75 MTPE score and a 2% usage ratio over a total of less than 1 million segments in the EN>NL language pair. Similarly, the big blue C in the middle of the chart details that the third NMT engine has an average 0.65 MTPE score and a 45% usage ratio over a total of more than 3.5 million segments for the EN>FR language pair.

Several insights came to light thanks to this scatter-plot. First, there is clearly an engine (engine C) that is used the most for each language combination, as nearly all its data points are placed to the rightmost part of the chart. However, if the corresponding MTPE average scores are considered, the most used NMT engine never coincides with the best performing one, for any language pair.

In addition, the best-performing NMT engines are the ones that are used the least. For instance, if we consider the EN>DE language combination, engine C is the most-used one (63%) but it has an MTPE average score (0.58) significantly lower than the score of engine A (0.72), which has been used only on 15% of the 2 million segments for that language pair, instead. This difference is even more

![Figure 2: NMT engine performance by domain and by language combination.](image)
relevant for the EN>NL language combination, where the worst performing NMT engine (engine C, average MTPE=0.64) is used quite often (90% usage ratio), while the best performing NMT engine (engine A, average MTPE=0.75) is almost unused (less than 3% usage ratio). It is worth pointing out that engine A is not available for the EN>CS language pair.

Further details are provided by the scatterplot in Figure 2, where NMT engine performances (evaluated as MTPE scores, on the vertical axis) are broken down in terms of source domain, per language combination. Source domains (horizontal axis) are reported with numerical IDs to make the chart more readable and their actual meaning is enlisted in the informative text area at the bottom right corner of the same chart (domain 0 stands for medical, domain 1 stands for travel and hospitality, and so on). Average MTPE scores per language combinations are shown as red reference lines. Each NMT engine is associated with a different color, so that several interesting elements can be spotted immediately.

First, engine A is the best performing NMT engine (in terms of MTPE) in the majority of source domains for the available language combinations (all except EN>CS). Second, the NMT engine C provides the lowest MTPE scores for the majority of source domains in all language combinations, since its data points are quite often below the average. Moreover, we can see that the best MTPE score has been achieved by NMT engine B (travel and hospitality, EN>NL) and the worst one by NMT Engine C (gaming, EN>CS). The NMT engine B scores on average for the majority of language pairs and domains.

**Conclusions**

Boosted by business globalization and supported by technological advancements, the MT market is gaining strength and importance, with significant growth rates expected in the next five years. Companies willing to improve productivity rates and to enlarge targeted customer segments cannot overlook MT solutions, as they promise a viable and affordable solution.

However, several challenges typically hamper the selection of the best MT engine: different MT approaches are available; multiple providers showcase heterogeneous service offerings; and MT quality differs significantly depending on applied MT technique, language combination and content domain. This often determines ineffective MT usage patterns by companies. These patterns were identified by analyzing MT usage data on three different NMT engines in four different language combinations and 11 domains. The MEMT approach can cope with these challenges, as it allows querying more than one MT engine at the same time, but it is not a cost-effective solution. From a business point of view, then, we can see that proper MT management systems are helpful in order to automatically identify the most suitable MT engine for a certain translation service request. [M]
Toward global machine translation

Arvi Hurskainen is professor emeritus at the University of Helsinki, Finland. He has been working on rule-based language technology since 1985, specializing in morphologically complex languages such as Swahili and Finnish, and has developed a number of applications in these languages.
Machine translation (MT) technology can be divided broadly into two very different approaches. The older technology relies on close integration of human knowledge about language on one hand, and on the calculation speed and power of the computer on the other hand. For a long time, this approach was considered the only feasible one. However, the advent of the internet and the increased power of computers made it possible to develop a very different approach to MT. The development led to various statistical approaches, where machine learning is key. The latest versions of the statistical approaches are known as neural machine translation. While it is difficult to fully understand how this system works, its development has currently usurped almost all resources in the study of machine translation.

The current situation is very unfortunate, because roughly 1% of the world’s languages fulfill even the minimal criteria for statistical or neural machine translation. This trend pushes the majority of languages into a marginal position, because if the language is not computerized, as the customary expression runs, it is not respected.

Most languages are doubly disadvantaged. The current trend in MT does not suit them, as they do not have the needed language resources such as parallel text corpora. The other disadvantage is that they are not commercially attractive, which means that market forces push them into the margins.

However, those languages can still be included into the language technology market. The rule-based approach to machine translation does not exclude any language. A language can be computationally described even on the basis of speech in any field situation without the internet.

**Grand aim of language technology**

The grand aim of language technology is to develop a system where two language users with no mutual language competence can communicate through an MT system. Ultimately, this system would join all people into a single communication network with no language competence barriers.

Currently we are still very far from this aim. The attempts to construct such multilingual translation networks often include mere tens of languages, and more languages are involved with various degrees of coverage and accuracy. There is often much enthusiasm among the system developers — but in this, they behave as though they’re the leaders of religious groups. Each group leader is convinced that his (they are often males) message is the only one that leads to salvation. These leaders go around to mission tours and try to persuade followers to their flock. I myself, having worked for decades with the language technology, including such languages as Finnish, English, Swahili and other Bantu languages, have been approached by these sect leaders, but so far with poor results. Not only do I have my own approach, I see this sect approach as unnecessary.

**Interlingua as mediator**

Why should we force all language technology developers into a single approach, if we can include all approaches into a joint global translation system? By using a clearly defined interlingua as a mediator between two languages, we can construct a global language translation system where no language and no approach is excluded.

The idea of using interlingua is in no way new. What is new is the use of a normalized version of English as interlingua, instead of Esperanto or other invented languages. This approach is also independent of translation methods. The only thing that matters is its capability to produce high-quality translation between the non-English language and this normalized English, to both directions. The good translation result may be produced by any system, including statistical and neural methods.

**Interlingua can have various levels of abstraction**

The big players in MT use the normal surface form of English as interlingua in multilingual translation. This is the simplest solution, because translation systems between a language and English are needed in any case. It is straightforward to use that version of English as interlingua.

The English as it is used today in media is far from ideal interlingua. It contains frequently occurring ambiguities, which are due to omitting clause boundary marks and even leaving off relative pronouns and other subclause markers. Also, the use of gerund form of verbs instead of infinitive forms causes unnecessary ambiguity.

When translation is done from a non-English language into English, these anomalies can be avoided. If the source language encodes such features, which would cause ambiguity in normal English, these codes can be used for avoiding ambiguity, when translating from interlingua into language C.

Consider the sentences below:

(a) *I would argue that the modified English described above is the minimum requirement for interlingua.*

(b) *I would argue, that the modified English, which was described above, is the minimum requirement for interlingua.*

The first sentence is written following standard English writing style. The latter is the same sentence translated from
Swahili to English. This version has commas and it also has relative pronouns. These seemingly small differences make a huge difference in machine translation, especially if we further translate the sentence into a third language such as Finnish.

Swahili requires that relative pronouns are overtly encoded and that clause boundaries are separated by commas. Not surprisingly, Finnish does the same. Why then should we lose all this critical information in interlingua by trying to imitate “fluent” English and then face huge ambiguity problems, which should be solved in any case?

When we translate sentences (a) and (b) above with Salama to Finnish, we get different translations.

(a) Minä väittäisin, että modifioitu englanti kuvaili ylempänä on vähimmäisvaatimus interlinguaan.

(b) Minä väittäisin, että modifioitu englanti, joka kuvailtiin ylempänä, on vähimmäisvaatimus interlinguaan.

In sentence (a) the information pertaining to the relative clause was lost, but in sentence (b) it was retained. Next, we test what happens when we translate the same sentence from Swahili to English.

Ningetoa hoja, kwamba Kiingereza kilichotoholewa, kilichoelzwa hapo juu, ni haja ya msingi kwa interlingua.

I would argue, that modified English, which was explained here above, is the basic requirement for interlingua.

The translation of the English sentence above to Finnish:

Minä väittäisin, että modifioitu englanti, joka selitettiin tassä ylempänä, on perusvaatimus interlinguaan.

We see that the translation from Swahili to Finnish via interlingua is better than the translation from standard English to Finnish. This is due to the fact that the structure of the sentence components is retained throughout the translation process.

Linguistic tags inherited from source language

In addition to using the modified English as described above, we have also other means for improving the translation process. When we use a rule-based approach in translation, we have a large array of linguistic tags, which can be carried on through the translation process.

This is not, however, a simple thing, because in order to translate from English into the third language, we need first to analyze the interlingua. Analyzers normally analyze only plain text, and tags attached to words disrupt the whole process.

It is, however, possible to construct an analyzer, which accepts surface words with tags attached to them. When the analysis is done, the codes can then be separated as individual codes. In the result, we will then have the codes inherited from the source text and the codes added by the new analyzer of English.

Normally the disambiguation of English is a very heavy process requiring a number of complex rules. Now we have a very different situation because we can compare the inherited tags and the tags inserted by the English analyzer. The rules needed for disambiguation in this approach are normally straightforward, requiring only checking the co-occurrence of old and new tags. Even syntactic mapping of English can be bypassed, because the inherited syntactic tags suit with minor modifications also to English text. The main points where modification is needed are the direction codes < > in tags to show the direction of the head. The word order of Swahili and English is fundamentally different.

For example, consider a sentence extracted from the World Machine Translation Challenge 2017:

Trump says he was being sarcastic when claiming Obama was the founder of ISIS.

The sentence has ambiguities that are difficult to resolve. The Swahili version of the sentence is:

Trump anasema, kwamba alikuuwa mwenye kejeli alipoda, kwamba Obama alikuuwa mwanzilishi wa ISIS.

Now we translate the sentence from Swahili to English, using a version of Salama, which attaches inherited linguistic tags to translated words.

Trump+n+hum+@subj says+v+pr:na+@fmainvtr-obj, that+conj+@cs he+pron+@subj was+v+past+@fmainvintr sarcastic+adj+@nadj when+conj+@advl he+pron+@subj claimed+v+past+@fmainvtr-obj, that+conj+@cs Obama+n+hum+@subj was+v+past+@fmainvintr the founder+n+hum+@p of+gen-con+@gcon Isis+@nh.

The reading contains tags for part-of-speech, TAM markers, syntax and semantics. This text form is then analyzed with such an English analyzer, which accepts codes attached to surface words. After analysis, codes are detached from the words, so that they can be handled using normal routines.
In the previous reading, we have two sets of codes: those inherited from Swahili (lower case) and those added by the English analyzer (upper case). The disambiguation of English text becomes now remarkably easier, when we have the encoding of both languages at hand. The reading where the encoding is the same is probably the correct choice.

Syntactic coding is inherited from the source text, and no new encoding is needed in this phase.

The text after light disambiguation is displayed below. Redundant codes are removed and the remaining codes are converted into upper-case.

"claim" v past @fmainvtr-obj V PAST
"claim" v past @fmainvtr-obj V EN
"<,>", COMMA
"<that>"
"that" conj @cs N Heur
"<*obama>
"obama" n hum @subj PROPN
"<was>"
"be" v past @fmainvintr V PAST SG3
"be" v past @fmainvintr V PAST SG1
"<the>"
"the" DET
"<founder>"
"founder" n hum @p N SG
"<of>"
"of" gen-con @gcon PREP
"<*isis @nh>
"isis" @nh PROPN
"<.$>".

. **CLB
Trump sanoo, että hän oli sarkastinen, kun hän väitti, että Obama oli Isisin perustaja.

**Semantic encoding**

Semantic tags inherited from the source text are especially valuable in translation, because semantic information usually cannot be inferred from the form of the words. Such semantic features as humanness, animacy, sex, time, place, proper names (male, female, title, organization) can be encoded in the analysis of source language and then be transferred to the interlingua and the third language. News texts of a local language have large numbers of such proper names, which are not known in English or in a third language. Inherited semantic tags help in translation, especially if such semantic distinctions have bearing on translation into the target language.

**Language-type-specific encoding**

If language A and language C have similar language structures, it sounds far-fetched to translate between them via interlingua. For example, two Bantu languages have normally almost identical noun class structures. It may be tempting to translate between them directly instead of using interlingua in between.
However, it is also possible to retain the noun class information of language A and use it when translating via interlingua to language C. A large part of nouns in language A and C have the same noun class category and they need no new encoding. If the class of a noun in language A is not the same in language C, such cases must be treated separately.

Also, linguistic gender is a dominant feature in many languages. However, there is seldom direct mapping of gender affiliation on a word level.

**Suitability of the system and capacity constraints**

It is evident that the proposed translation system favors rule-based approaches, because they encode the linguistic information comprehensively and accurately. Statistical and neural translation systems do not do that. And if they try to make linguistic information explicit, the coverage and reliability is so low, that it is of little use in translation. However, if no modification of the interlingua is expected, those systems also work.

A comprehensive MT system often requires a large number of sequenced processes and handling of tens of thousands of rules. When we add to this the second translation round from interlingua to language C, much more computing is needed. However, the capacity requirements depend very much on the comprehensiveness of the translation system. If the vocabulary is limited to the vocabulary used in prose texts and conversation, capacity problems do not occur. Also, domain-specific approaches can be used for reducing capacity requirements.

**Conclusion**

I have suggested that a global MT system can be constructed using an open approach where emphasis is on high-quality translation between a non-English language and interlingua, which is a modified version of English. I have demonstrated the feasibility of this approach by testing it between two languages, Swahili and Finnish, both morphologically complex languages. In addition to using the modified version of English as interlingua, it is also possible to use a selection of linguistic tags of the source language to aid in further translation process. The system allows any translation method to be included into the system. The main requirement is the translation result of high quality between the local language and interlingua.
Modern tech meets Indian Sign Language

Ribhu is a freelance journalist based in Bangalore. He is fond of traveling and covers tech, arts and culture, environment and wildlife.
Indian Sign Language (ISL), also known as Indo-Pakistani Sign Language (IPSL), is one of the most common sign languages in South Asia, used by millions of speakers in India, Pakistan and Bangladesh. However, as is the case with some sign languages, it is difficult to accurately estimate the number of people who practice ISL. The study on sign languages in India meets a massive roadblock in particular, as the actual number of deaf people has not been ascertained. This barrier creates an enormous problem for individuals and organizations who want to have new technologies implemented around the usage of sign languages in public spheres.

The 2011 Indian census stated that roughly 1.3 million people in India experience hearing impairment. These numbers are quite different from that of India’s National Association of the Deaf, which puts this number at 18 million, or roughly 1% of India’s population. But even these numbers are low when they are compared to global standards — it has been reported that 5% of the world’s population and 3.5% of Americans experience hearing loss.

The pattern of underreporting is not surprising when it is compared to other forms of disability. The 2011 Indian census reported that 2.21% of India’s population is disabled, while the global average for that year stood at 15%

The reason for such underreporting is partially due to the social stigma around disability. This means census takers experience an inability to accurately report the cases of disability. The resulting data then can impair the formation of an effective roadmap for the implementation of sign languages in universities and other public spheres.

A rudimentary stage

Unlike the sign languages in the west, ISL is at a rudimentary stage, where it is still struggling to be acknowledged as a language for minorities. Acknowledgement would help it garner more investment of resources so that it could expand and adapt to modern platforms.

Even though ISL is used by numerous deaf people in India, it is still not an integral part of the teaching curriculum. It was only as recently as 2005 that the National Curriculum Framework (NCF) in India acknowledged sign languages and hinted at them qualifying as an optional choice of language for the hearing-impaired. Since then, numerous independent and organizational efforts have been made at implementing sign languages in different curriculums for different languages, including a chapter on sign languages in by NCERT in a third-grade textbook.

ISL is unrelated to either Hindi or English, although it draws some signs from British Sign Language and its alphabet is based on English. It has its own dialects, too. “Indian Sign Language is very scientific and has its own grammar, but lack of awareness has meant that many deaf people are not even aware of institutions where they can learn it and equip themselves for public communication,” said Andesha Mangala, assistant professor at the government-funded Indian Sign Language Research and Training Centre (ISLRTC) in a BBC report.

The need for a comprehensive system around sign language starts with an official dictionary for it. The ISL dictionary (now upgraded to a second edition) was launched in March 2018. Presently, it contains 6,000 words, in categories such as academic, medical, technical, legal and everyday terms.

However, the presence of a dictionary certainly doesn’t put an end to the challenges. There are not enough schools in India that cater to the needs of disabled students, and the ones that exist face problems with funding and expensive technologies that are imported from other countries.

Retired colonel Kamrinder Singh is the secretary of the Society for Welfare of the Handicapped, which was established in 1967, and manages three schools — Patiala School for the Deaf, Patiala School for the Blind and Patiala School for the Deaf-blind — in Saidipur village near the city of Patiala in Punjab. Currently, there are 400 students in these NGO-run schools.

“There haven’t been any technologies that cater well to the needs of our students in India, and we have been importing most of our technology from countries like USA, Japan and Germany, and they tend to be expensive. The Indian versions are free of cost, but they are not up to the mark compared to their western counterparts,” said the retired colonel.

Singh hopes that the software gets better with time, as it would add to the cost-effectiveness of these programs that are not funded by the government. “Disability is a state subject, and more often than not, these roles are not played well due to the lack of funding,” he added.

Although the launch of the dictionary has opened up doorways to mass-implementation of ISL on primary fronts, the main challenge still is the absence of easily accessible tech-based solutions around sign languages that could work alongside human interpreters. India’s deaf population is largely illiterate, so even reading public signs can be a challenge.

However, home-grown solutions on these fronts are emerging.

Researchers at the Punjabi University in Patiala, India, have developed an automatic translation system based on ISL that takes English text as input. The program then
Around four years ago, I was visiting the school in Saidipur village in Patiala, and I came across a lot of students who are hard of hearing. This is when the idea came to me. I thought that we need to have a system through which they can read the texts in their curriculum — a simple interface that would allow them to copy things to the interface and have them converted to a sign language,” said Vishal Goyal, who has been working on this project with the aid of his colleague Lalit Goyal and students at Punjabi University.

Vishal Goyal has worked extensively in machine translation (MT) systems for Indian languages, including developing a Hindi to Punjabi MT system copyrighted in 2011 and available for use at http://h2p.learn-punjabi.org. “Earlier my research was in natural language processing, and this idea came to my mind that since machine translation is my research area, I should use this system for the community,” he said.

How the program works

The program, which was copyrighted in 2018 and is in its final stages of development, uses an English parser for parsing texts. The input in English is broken down to get the phrases, structure and grammar of the sentence. After that, the phrase-reordering module reorders the words of English sentence according to the rules or ISL grammar.

Then, the eliminator module of the program gets rid of unwanted words from a sentence such as linking words, articles and so on. The output of the eliminator module is then sent to the lemmatization module, which converts the words into the root form. Now, each root word is checked for availability in the English ISL dictionary. The unknown words in the sentences are replaced with their synonyms. If a synonym is not available, the fingerspelling of the words is performed character by character. The words are then played out through an animation tool.

This application requires less memory as compared to human sign language videos, and is highly flexible and programmable as it can be altered using synthetic animation of various kinds — animated kids, grownups or animals.

This program happens to be the world’s first working system for ISL, and the scope for its applications are huge, especially in the public spheres.

Goyal and his team are pushing for the program to be implemented at a national level through various organizations operating in the field. For instance, if the program is implemented through the University Grants Commission (UGC) in Indian universities, it can cater to the needs of students who are hard of hearing. The application of such program is not just limited to universities, but also at public places such as hospitals, airports and railway stations in India that do not yet have a functional system to cater to the needs of hearing-impaired people.

This system stands out from the past work that has been done in this field as it has more animations along with a better flow in them. It also has more vocabulary which makes the program more efficient.

A promising scope

This work in the field of sign languages has started seeing fruitful results. The government of India has approved a pilot project using ISL with the collaboration of Punjabi University. It has already asked the board for a proposal which includes the scope, utility, specifications and expenditures for the project, which is to be piloted at the New Delhi railway station.

The project will serve as a model for other railway stations in the future. Henceforth, it is expected that the program will garner interest from other organizations that operate in the public sector.

Once the program gets more funding from government bodies, it will be capable of working with not just English, but also other Indian languages to provide more regional context in the areas of its operation.

For his contributions to the field, Vishal Goyal received a state award on World Disability Day, December 3, 2019. Presently, he is overseeing
other sub-projects that can further extend the capabilities of the program, including an automatic telecast of news in ISL through synthetic animations; a machine translation system from Punjabi; text to ISL synthetic animations; and translating complex and compound English text sentences to ISL synthetic animations. Along with his team, he is also working on a similar model which could be implemented for airport announcements. This program has already been funded by a private organization, and if things go well, it could be implemented in airports.

However, the biggest challenge still lies with the acceptance of this program with core organizations that are responsible for standardizing the ISL, such as the ISLRTC, which was set up by the government of India 2015.

“The biggest challenge is to work together with the ISLRTC because they have the people and resources to test this system on the ground. It will be great to work with their resources and expertise,” said Vishal Goyal.

The Ministry of Social Justice and Empowerment, which oversees ISLRTC, said that it is open to new ideas around technologies that can help benefit people with disabilities. “If someone wants a collaborative effort with the government, they can write to the secretary with a detailed plan,” said Mrityunjay Jha, the deputy secretary at the Department of Empowerment of Persons with Disabilities, operating under the umbrella of the Ministry of Social Justice and Empowerment.

“I gave a demonstration of the software to ISLRTC a few years ago. Even though they liked our work, their only concern was with introducing facial expressions, that they were not up to the mark. However, if they show interest in a project like this, this is something that we can improvise on as per their needs, as the foundation for such a system is set,” explained Lalit Goyal when asked about working with ISLRTC to expand the applications of the program.

There has also been parallel work done to have a mobile application which contains a video catalog of all the words released by ISLRTC. The app, simply named SignLearn - ISL learning app, is available on the Google Play Store for free download and has already been downloaded over 5,000 times. It is based on the video catalog that is released by ISLRTC.

Jestin Joy, the developer of the app who has been working on it as a part of his research for around five years, thinks that such a platform can be helpful in communicating with hearing impaired individuals, as it has a vast catalog of around 4,500 signs.

Since learning sign language requires a lot of hand movements, learning about it through text books is not easy, and this is where multimedia platforms can come in handy.

“This is a dictionary for those who are hard of hearing. If you have a spoken language, then a normal dictionary is okay, but we need to describe words as signs and fingers movements, and this is where multimedia solutions will be helpful,” Joy said.

When asked about the implementation of this synthetic animation program to modern platforms such as mobile apps, Vishal Goyal seemed quite positive. “That’s not a big deal. Once we gain some traction and acceptance by more organizations in India, these are minor things that can easily be taken care of,” he said. [M]
When internationalization isn’t enough

Arle Lommel is a senior analyst for Common Sense Advisory (CSA Research).
In June 2019, Politico reported on the travails of Epic, a leading US maker of digital health software, in adapting its product to the requirements of the Danish health system. The company discovered that translating the user interface was far from enough. Although some of the problems Danish healthcare workers faced were the result of subpar translation, others were more fundamental. For example, a lot of Epic’s functions were around customer billing, but in the Danish single-payer system, these features were irrelevant. Similarly, the system had hard-coded roles for nurses and doctors based on US legal requirements that prevented them from doing their job in Denmark. One physician described the results as “indescribable, total chaos” and Politico reports that the same physician found that “Epic might work in the United States [...] but its design was so hard-coded in US medical culture that it couldn’t be disentangled.” After the Danish government spent over $500 million over three years, the system still is not fully functional.

Why should such a system struggle? After all, the art of internationalization has been clear for the better part of three decades: you write a culturally and linguistically neutral code base, to which you apply language as a “skin” that can be swapped at will. In this model, the goal is to remove all assumptions from the underlying logic of a program and create something that can be deployed anywhere from Baghdad to Wuhan. Although software often fell short of the idea — particularly when dealing with “difficult” writing systems — the methods and approaches are well understood.

In the internationalization paradigm, as shown in Figure 1, developers build a code base of core technical and business functions that they ruthlessly scrubbed of all linguistic and cultural references. Strings, colors, icons, dates, times and anything else that would vary from country to country had to be “externalized”: placed in separate files where it was easy to change them as needed. In some cases, data contained culturally specific information, such as in the case of addresses, where no single model can capture all of the variants. Developers would then build complex structures to represent the variety of forms in the world and store abstract pointers to the data in the core product.

On top of this internationalized core sat multiple localized user interfaces. These determined what fields and data were displayed in the case of complex structures such as...
as addresses. The results might vary along both linguistic and geographical lines. For example, a German user interface in Germany might use euros for currency and spell words with the character ß but Switzerland would use Swiss francs, and ss instead of ß. Nevertheless, despite such variation, the assumption was that one code base could serve the entire world with a few adaptations here and there.

So why did Epic struggle? Indications are that the company did internationalize the code properly. Even if it did not, the problems the Danish healthcare system found were not something internationalization could solve: internationalization works well when software functions are themselves truly neutral with respect to culture — such as a tool for making technical drawings or designing a database — but breaks down as software becomes more complex and interacts with culture, custom, and language in more complex ways. For some cases, such as supporting a new writing system, it may be enough to add additional features that apply only to particular languages — but in other cases no single code base can support all requirements.

This was the situation in which Epic found itself. Its basic functions could not be written in a culturally neutral manner. Such problems become almost inevitable as software becomes more and more complex and ties together more and more knowledge and functions. Even something as simple as a function to send a package must deal with the realities of hundreds of different postal systems, not to mention customs, tariffs, taxes and regulations. When faced with vast number of variants of requirements both known and unknown, seemingly reasonable and innocent assumptions can lead to unintended consequences. Hard-coded database structures, processes, sequences and requirements can complicate translation and adaptation. In some cases, comparable functions may be just different enough that no amount of translation will make them work.

Add in the current trends to build functions based on machine learning, and the problem becomes more severe. Few organizations have training data for anything beyond their source language, usually English or Chinese. Although it is tempting to slap machine translation on top of English training data, this approach runs the risk of replicating American approaches to complex tasks that do not apply to other regions. The alternative, data manufacture, is expensive and can lead to poor outcomes if enterprises do not anticipate local needs accurately. In other cases, lack

Figure 2: Complex applications require locale-specific functions.
of data can lead to product failure, such as in the case of facial recognition functions that do not recognize certain ethnicities because they were trained on data sets representative of specific countries.

So what can organizations do when their software must interact with complex business or cultural requirements that vary from country to country? The key is to capture the variations in functions and processes from the beginning, to internationalize what can be, and to modularize your code base to use different modules when needed. As Figure 2 shows, much of some application may consist of local versions of code.

Critically, the local versions of modules should themselves be internationalized. For example, a Germany-specific tax module should not assume all implementers will speak German. Increasing geographic mobility means that expatriate operators may need to see the locally relevant information in English, French, Chinese or any other language. Even though the functionality may be specific to Germany, the localization layer needs to support any language. You may not localize each version into multiple languages, but it needs to be localizable. However, the degree of internationalization may vary. For example, although you might translate the Germany module, you would be unlikely to need to use different calendars or non-euro currency, so you might not internationalize those aspects.

One of the challenges of creating such complex products and determining what each module must support is that core software development teams seldom consult with localization teams or local implementers. As a result, they may set certain decisions in stone before they understand the implications. At that point, proper localization may require them to rip and replace significant portions of code and then reengineer solutions for their markets, even if they have properly internationalized the code.

The solution in such cases is to bring together a team to define requirements early on, and determine what components can be localized as-is, which ones need to exist in multiple versions and what training data may be required. This team needs to include experts in regulatory and legal compliance, local business practices, data requirements, local languages, finance and any special subject areas of concern. These individuals do not need to be involved at every step but should sign off on initial plans and any substantial changes. Involve them again in early-stage reviews and be sure to include experts not just for current markets and languages, but also for any you expect to add in the foreseeable future.

This approach does not mean you need to build everything all at once, but rather that your development should plan for international growth and requirements. For example, even if you will not initially need multiple tax modules, building in the ability to switch them — rather than hard-coding them — will make it easier to adapt your product in the future. In such cases, you can start by identifying every point where your software interacts with laws, customs or business practices that may vary, and modularize your code around those points. This lower-cost approach also makes it easier for you to deal with any changes that may occur in your home market without needing to reengineer your products, and so may save money in the long run, even for your home market.

You can achieve success in building complex software applications that rely on algorithms or data if you think beyond the simple internationalization paradigm that has dominated software development for the past 30 years. Treating international markets seriously will make your products more relevant and useful and increase brand loyalty. The up-front costs may seem high, but they will be small compared to the costs of market failure or reengineering the core of a product after launch. This is a case where an ounce of planning and prevention can save pounds of trouble later on.
How multilingual chatbots influence localization

Kaspars Kaulins is a business development director at Tilde, a European language technology innovator and service provider offering translation, localization and custom machine translation.
As technology advances, users are becoming more demanding and it is becoming harder to meet instant service expectations from customers. Companies also face the challenge of multilingualism as they are pressured to bridge language barriers and provide content in the customer’s native language. Thankfully, AI and machine learning (ML) have created opportunities for companies to deliver a consistent user experience across languages by using AI-enabled, intelligent virtual assistants. Deep machine learning and language data are two major components that can make or break these human-like interactions. Chatbots not only enable companies to serve their customers 24/7, but also create an exciting and emerging field for the localization industry.

Evolution of chatbots

We live in an era where chatbots are springing up like mushrooms after a rain. While the first chatbot was released in the 1960s, it is only in the last couple of years that this chatbot generation has reached its current state, where it can converse in a manner that feels nearly human. The advances in AI and ML, and the revolution in natural language processing (NLP) and natural language understanding (NLU), have created a generation of chatbots so advanced that often people don’t even realize they are communicating with a machine. The chatbot revolution has had an impact on all of us — from the way we live and work, to the way we interact with one another. AI has powered virtual assistants to handle complex human interaction with ease. They can help serve customers on websites, online stores and helpdesks, and automate the handling of basic user requests.

According to Gartner statistics on chatbots, experts predict that by 2021, 85% of customer interactions with companies will be handled through chatbots. Virtual assistants provide clients with instant answers and allow companies to reduce costs. Best of all, human agents don’t have to answer the same question 100 times a day, so they can spend their time solving more complex issues. The online publication Chatbots Magazine predicts that the chatbot market will continue to grow — $5 billion will be invested in chatbots by 2021. This will allow companies to cut operational costs by up to 30%. According to Juniper Research, by 2023, the use of chatbots will save both businesses and consumers 2.5 billion hours. When time is a scarce commodity, saving so much time for both businesses and customers is outstanding.

Many companies have already integrated chatbot technology, and numerous others plan to do so in the near future. No matter whether this technology is known or new, companies all face the same challenge: creating human-like assistants that provide a universal customer experience across languages. This adds an emerging field to our industry: chatbot localization.

Incorporating linguists into development

One of the most important elements that makes up a chatbot is language and its ability to converse, yet not all chatbots are created equal. A virtual assistant’s knowledge-based texts, their quality or lack of it, and data processing can make or break the customer experience. Chatbot developers not only need to invest resources in technology, but also in chatbot language development. The virtual assistant knowledge database consists of two types of texts: questions and answers.

Knowledge database development starts with crafting potential customer questions. A well-established question database gives the virtual assistant direct insight into the client’s needs and issues. Most customer service teams deal with up to 80% of the same frequently asked questions each day. These are repetitive, low-level questions that could be easily answered by an AI powered chatbot. Over time, these customer service teams have accumulated decent baseline data regarding customer needs and wishes. To train a virtual assistant to understand client needs and provide human-like interaction, it needs enough diverse training material that reflects the way customers talk and includes the many ways in which clients might express the same thought.

For example, a customer seeking help from an online banking chatbot may ask, “Where can I pay?” while another client might ask the same question in the form of a statement, such as “I want to pay” or “I need to make a payment.” To be able to process a client’s needs properly, virtual assistants need to be trained in every possible way of asking the same question. The virtual assistant’s knowledge expands further through use, as its database is constantly reviewed and improved, and through the addition of new data obtained through interactions with customers.

To compile a good knowledge database, linguists need to use their creativity and have sufficient knowledge of the customer persona in order to really fit into the skin of a regular customer. Sometimes, people don’t have a clear structure in the way they talk, and they often make spelling mistakes or grammatical errors when they are writing. The inconsistent use of different word forms, unjustified changes in expression, grammar and spelling errors can be confusing for virtual assistants and can disrupt their ability to converse if they are not trained for such situations. Therefore, creators of virtual assistants are encouraged to hire localization specialists, text editors and proofreaders to properly prepare virtual assistants for human-like interactions. For the sake of better understanding, texts that
include typos and grammatical errors need to be included in the virtual assistant’s knowledge database.

The other facet of intellectual virtual assistants involves the responses that the client receives from the chatbot. To create human-like answers and converse with clients, it is crucial to develop responses in a situationally-appropriate language style, which requires additional effort from language editors as well as proofreaders and translators. Experience has shown that copying plain text from business or government websites, which frequently does have answers to a user’s question, is not a good solution for building a chatbot knowledge base. The purpose of having a virtual assistant is to provide customers with a pleasant experience. Having a virtual assistant is not enough; it must also be intelligent, with high-grade quality and the ability to adapt to different scenarios and situations.

**Chatbot personality and tone**

A virtual assistant’s goal is to provide the customer with short, quick and easy-to-read responses that offer an immediate and fitting solution. At the same time, it is important to consider and establish a clear chatbot identity. The language of the virtual assistant should be in line with brand values and tone of voice of the company and must clearly express its defined personality.

Depending on the target audience, companies must choose one consistent conversation method that could be either familiar or formal, scientific, professional or conversational, or they could define their own unique dialogue style. If the virtual assistant is expected to be kind, it should use courtesy phrases like “please” and “thank you.” If it is intended to break down the bureaucratic barrier between the institution and the client, the characteristics of a scientific and business language style must be dropped. One of the challenges that comes with playful chatbot personalities used in sales is the use of emojis and gifs. Not only do they create additional technical challenges regarding compliance with computer-aided translation tools, but they also require translators to be fluent in emojis and pop culture references so that conversations do not lose a personal touch or context.

**Mixing machine and human translation**

Machine translation (MT) can benefit companies that want to make their virtual assistant multilingual as MT can significantly speed up processes. However, it is best practice to involve a linguist who will post-edit translations for cultural references, language nuances like jokes and aphorisms, and make sure that the original text expresses the same goal as the source language. Incorrect translations can cause companies to lose potential clients, and nobody wants to deal with situations which could be easily prevented.

Often a linguist’s daily workflow is quite repetitive, and they may fall into unwanted routines. All that changed recently, when new and exciting opportunities presented themselves through chatbot localization projects. One of our linguists with more than 15 years of experience in the industry shared her insights of working on a project for Nuance, a pioneer and leader in conversational AI innovations. She also reminisced about the old days when editors used to proofread translations by printing out documents and using a red pen. At that time, she could never have imagined working with texts for virtual assistants. However, this is the beauty of technology — innovative ideas can be turned into reality, which requires an open mind and flexibility. Translators and editors need to be able to adapt and quickly and to take on new tasks and challenges. Nowadays, more and more translation projects involve dealing with AI, whether it is post-editing MT translated texts, localizing chatbots or proofreading texts generated by voice-to-text technologies.

A chatbot localization project is challenging, but also very rewarding. At first, it is quite difficult to adapt. You go from translating and editing medical or legal texts to thinking about all the possible ways how one question can be asked and answered. You need to use your creative side when crafting texts for virtual assistants, and you must understand the industry background. There are two components to chatbot localization. Firstly, creating scenarios in the source language and secondly making the chatbot multilingual.

For the Nuance project, chatbot scenarios were first created in English and then localized into the three Baltic languages. For example, there are some questions that are formally included in Latvian scenarios. However, because of the market-specific product offerings, they are not used since the products and services available in Latvia might be different to those on offer in Estonia or Lithuania.

That is the added value of involving a localization specialist as opposed to plain translation. Localization specialists consider language nuances, cultural differences, traditions and regional specificities. You can’t just translate a text directly without thinking about all the small nuances — formal word-for-word translations just don’t work when localizing chatbots. Localization specialists must use their experience to make sure that the chatbot greets customers and answers their questions in a way that is familiar and recognizable in domestic markets.

The beginning of this project was very hands-on, with endless hours per week spent with the client to create appropriate texts. As the chatbot project moved into production, the workload diminished. However, work to keep improving the virtual assistant is constant. Linguists need to review the content to see how people have interacted with the chatbots and adapt it accordingly by editing texts, adding new options and moving incorrectly answered questions to the right answer groups.
An exciting part of such projects is the ability to see the result of your work by literally interacting with a virtual assistant whose language you have helped to shape. Many times, translators return a completed project back to the client and they never see where the translation is used, but now they are literally able to interact with texts they have created. Overall, the experience of working on chatbot localization is challenging, exciting and rewarding.

**Insights from an egov chatbot in Europe**

Virtual assistants are mostly seen in customer service, online stores, websites and global companies; however, the public sector is also adapting these AI-driven technologies. UNA, one of the first intelligent virtual assistants in the public sector, was launched in 2018 and it was the first public administration virtual assistant in Latvia. UNA was the Register of Enterprises’ virtual employee, and this proved that virtual assistants can not only benefit enterprises, but can also help government institutions. Since then, we have developed over ten virtual assistants for various public administration institutions. We are also working on creating a unified platform for all the virtual assistants serving the public administration sector in Latvia that will be launched this year.

UNA helps clients register enterprises, liquidate companies and check documents, so human customer support can focus on the creative problem-solving that it does best. UNA has been created and trained to understand questions and respond with natural language in both Latvian and English. It incorporates multilingual NLP and NLU services for knowledge training. UNA can solve about 30% of client inquiries by itself. More complex inquiries that can’t be handled by a virtual assistant alone are passed onto the appropriate customer service representative.

A common concern is that AI might take away people’s jobs, which is why we involved the people from the Register of Enterprises to help in the development of UNA. Not only did they provide invaluable insight to make UNA better, but they also found that UNA will make their job more pleasant, instead of making it redundant. Over 64% of business respondents surveyed by Statistica believe that chatbots will make their work more enjoyable, instead of making it redundant. Over 64% of business respondents surveyed by Statistica believe that chatbots will make their job more pleasant, instead of making it redundant.

The future is unknown

More and more companies recognize the need for intellectual virtual assistants that can provide immediate customer support. Moreover, they see the increased value of conversational AI that can interact with the customer in his or her native language. We can all agree that chatbots are here to stay, and that they are getting increasingly intellectual and savvy. Virtual assistants will become multilingual not only in text, but also in voice, which will create a whole new set of challenges for the audio localization workflow that will affect both companies and localization specialists.

To provide the most human-like conversational AI experience, we need to include human intellect (talented translators, linguists, subject-matter experts and copyrighters) as a core value for the development of both text-based and voice-based virtual assistants. Localization of chatbot dialogue scenarios and knowledgebase requires a specific approach and skill sets, and therefore translation companies will need to develop respective competencies and capabilities. Given the growing demand and business opportunities in various industries and sectors, this is an effort well worth making.
Web globalization lessons from Wikipedia

John Yunker is author of *Think Outside the Country: A Guide to Going Global* and *Succeeding in the Translation Economy* (available in both English and Japanese editions). He is cofounder of Byte Level Research and author of the annual Web Globalization Report Card.
Two decades ago, I began studying the globalization of websites. At the time, the practice was — like the internet itself — in its infancy and there were few established best practices. But by talking to the people who managed global websites and managing localization projects myself, I began to learn which practices worked and which did not. This led to the formation of an annual benchmarking report known as the Web Globalization Report Card. Now in its 16th edition, this ranking of the leading global websites highlights those websites that excel across a number of categories, including global reach, consistency and depth of localization.

This year, Wikipedia emerged number one overall. As one of the world’s most-popular websites, Wikipedia relies on the contributions of volunteer writers and editors from around the world in creating and translating billions of words across millions of articles. And while Wikipedia may stand alone in many ways, it exhibits a number of web globalization best practices that any organization can benefit from. There are five key lessons one can take away from Wikipedia’s global success.

**Lesson #1: When it comes to languages, there is no finish line**

As I’ve said for many years, the internet connects devices, but language connects people. Companies that are serious about going global must be equally serious about supporting languages. Which languages you support will ultimately be determined by your global strategy, but if you were to follow the lead of the world’s global brands, you might aim for 30 languages, at a minimum.

Shown in Figure 1 is the average number of languages supported by the leading global websites according to the 2020 Web Globalization Report Card. Since 2004, this average has more than doubled to 33 languages today.

This average is based on studying the leading global brands, companies like Apple, Coca-Cola, BMW and Toyota. If we were to include a much larger sampling of websites, the average would most definitely drop. But one fact holds...
true — the steady, inevitable growth in languages across all websites in all industries. As more people around the world go online, the language requirements an organization must meet to communicate with them will only increase.

There was a time when English would connect you with the majority of internet users, but those days are long behind us. Figure 2 illustrates the leading languages spoken by the world’s roughly 4.5 billion internet users. Chinese is now the dominant online language, followed by English. But even if you support these two major languages, you’ll only reach 40% of all internet users.

The largest slice of this pie chart is “all other languages.” This slice includes languages such as Italian,
Dutch, Norwegian, Finnish, Swedish and many other languages that any global website probably already supports. What this means is that you can support the top 15 languages on the internet and still miss vast numbers of potential customers.

Support 48 or more languages, and your website will communicate with approximately 95% of the world’s internet users — which is the ultimate reason why companies such as Google and Facebook surpassed 48 languages years ago.

Which brings us to Wikipedia, with support for more than 280 languages. Wikipedia is one of the world’s most multilingual websites, and a testament to just how many languages the world’s internet users want to see supported, as they themselves created this content. Yet even Wikipedia is far from complete in its linguistic journey; its language reach will extend as high as its user base wishes it to go. When executives ask me how many languages I think they should support, I often point to Wikipedia as evidence of what the world’s internet users want to see supported.

So even if you’re just starting out with a handful of languages today, prepare for a journey that will no doubt see this handful of languages expand over time. Web globalization is a road without end, and successful organizations understand how to
design websites to support a diverse range of languages and scripts.

**Lesson #2: Global templates enable efficient globalization**

Your company may be only focused on expanding into one country today, but three years from now you may have 25 country websites to manage. In order to efficiently manage all of these websites, you need a global template.

Wikipedia could not successfully support so many languages across millions of articles without global templates. Figure 3 shows an article on the Pacific Madrone tree in English and Russian, both utilizing the same template. Should a volunteer editor wish to create this article in another language, this template will simply be reused.

Global templates allow regional and local teams to focus on content rather than design, a much better use of their time. And global templates provide a more consistent user experience across your many geographic websites and brands. Visitors may arrive at your global .com site and then navigate to a local website; if the designs change significantly, this may create an unneeded distraction.

Global templates may also be successfully leveraged across multiple brands. Figure 4 shows three Microsoft products — Microsoft, Windows and Office — all sharing essentially the same template. Given the benefits to global templates, one may wonder why all companies don’t support them. The major reason is that these companies are often quite decentralized, with independent regional and local offices that are not only comfortable maintaining independent websites but consistent on remaining that way. For those in the global office, it can be difficult to convince local offices to embrace global templates, but the effort is well worth it. Global templates allow organizations to be globally consistent while remaining locally flexible.

**Lesson #3: Think mobile when thinking global**

Most of the world’s internet users access the internet through their mobile devices. So it’s critical that the global templates you develop not only adapt to mobile devices, but are optimized for mobile users. The Wikipedia web page easily adapts its content to mobile and desktop browsers.

Adapting to the device is step one. Adapting to the user and their mobile needs and requirements is much more important. For instance, if the person understands only Russian, you don’t want to send English-only content.

Fortunately, Wikipedia uses language negotiation (also known as language detection) to greet users with the language most closely aligned with the user’s web browser setting. Language detection is one of four elements that go into forming an effective global gateway strategy, which also includes country codes, geolocation and a visual global gateway menu. These elements vary by organization but all can play a critical role in helping visitors quickly and seamlessly find their local content.

And what about offering users a mobile app? Wikipedia offers one as well, which features an interface localized into more than 50 languages, far more languages than most other mobile apps currently available.

What’s interesting to note about the app is how it makes navigating across languages within a given article easy to do. By clicking the languages button circled in red (Figure 5), a user can switch to a different language, which is a nice feature for multilingual users in search of all available content.

**Lesson #4: Keep it “light” to improve the user experience**

As companies expand their global reach, they inevitably discover that not all mobile subscribers enjoy unlimited data plans, let alone the buying power to pay for unlimited data. They also discover that many regions around the world are still
dominated by slower networks and older devices. Which means that those companies that localize their websites to remain fast-loading and functional for users with older devices and slower connections are going to be perceived as more user friendly within these markets.

For a typical mobile web user in India, a web page that “weighs” 8 megabytes (MB) a page could take 20 seconds or longer to display. I mention 8 MBs because, according to the 2020 Web Globalization Report Card, the average weight of a mobile home page has more than doubled from 3.1 MBs in 2016 to 8.3 MBs today. This is a troubling trend.

Except for Wikipedia.

While most websites reviewed for the Report Card weigh well in excess of 1 MB, Wikipedia is one of the few websites to weigh less than 1 MB. The Wikipedia home page weighs less than the home pages for Google and Facebook, two companies obsessed with performance. One way that Wikipedia keeps its website so lightweight is the absence of scripts that track users and usage, something no corporate website would think of going without.

I recommend that all websites target a ceiling of 1 MB — an aggressive goal to be sure, but one that will ultimately benefit your customers and keep you ahead of competitors. Limitations can be highly positive because they force everyone to do more with less.

Lesson #5: In the end, it comes down to relevant content

Like so many corporate websites, the majority of content on Wikipedia is available in English — more than six million articles. German comes in second with support for more than 2.4 million articles, followed closely by French. But when it comes to less-used languages such as Gaeilge or Kurdish, you’ll find 10,000 or so articles, or fewer.

But just because you don’t offer parity with English on your website doesn’t mean you can’t create a valuable experience for visitors. As shown in Figure 6, Wikipedia highlights links to timely and newsworthy content.

This lesson carries over to all organizations. Just because you don’t offer all content across all languages doesn’t mean you still can’t offer useful and relevant content to users. But you need to have a plan in place for managing language and content expectations. By all means, do not make the mistake that many companies make by launching what I call “local façades,” in which only the home page and a few subpages are translated. Local façades create negative first impressions; visitors quickly discover how little is available to them and are reluctant to return, which could do far more harm than good in the long run.

And Wikipedia is well aware of the content and linguistic gaps. It recently acknowledged that it needs to invest in helping users in emerging markets such as the “Global South” to more easily create and edit content using mobile phones on limited-bandwidth networks. In the end, content is what will make your website, and any website, worth visiting.

Wikipedia remains one of the world’s most worthwhile places to visit. I find I land there a few times a week, and it’s almost always time well spent.

Figure 6: Wikipedia’s Catalan homepage.
**Associations**

**elia.**

*European Language Industry Association (Elia)*

Elia is the European not-for-profit association of language service companies with a mission to accelerate our members’ business success. We do this by creating events and initiatives that anticipate and serve our members’ needs in building strong, sustainable companies, thereby strengthening the wider industry. Elia was founded in 2005 and has since established itself as the leading trade association for the language services industry in Europe.

Elia Brussels, Belgium  
Email: info@elia-association.org  
Web: http://elia-association.org

**Globalization and Localization Association**

The Globalization and Localization Association (GALA) is a global, nonprofit trade association for the language industry. As a membership organization, we support our member companies and the language sector by creating communities, championing standards, sharing knowledge and advancing technology.

Globalization and Localization Association  
Seattle, WA USA  
+1-206-494-4686  
Email: info@gala-global.org, Web: www.gala-global.org

**Conferences**

**LocWorld**

LocWorld conferences are dedicated to the language and localization industries. Our constituents are the people responsible for communicating across the boundaries of language and culture in the global marketplace. International product and marketing managers participate in LocWorld from all sectors and all geographies to meet language service and technology providers and to network with their peers. Hands-on practitioners come to share their knowledge and experience and to learn from others. See our website for details on upcoming and past conferences.

Localization World, Ltd. Sandpoint, ID USA  
208-263-8178  
Email: info@locworld.com  
Web: https://locworld.com

**Desktop Publishing**

**Global DTP**

Global DTP s.r.o., based in the Czech Republic, offers professional multilingual desktop publishing and media engineering solutions to the localization industry. Over the past 15 years, Global DTP has become one of the leading DTP/multimedia companies. We have been delivering high-quality and cost-effective services for at least eight of the top 20 LSPs and many other companies/agencies. Due to our extensive experience in localization and knowledge of the prepress, media and publishing industries, our team of 20 in-house professionals handles more than 1,000 projects every year. Our core services are multilingual desktop publishing, multimedia and eLearning engineering.

Global DTP s.r.o. Brno, Czech Republic  
+420 3 574 709  
Email: info@global-dtp.com  
Web: www.global-dtp.com

**Hornet Design Studio**

Since 2005, Hornet Design Studio has been focusing on delivering quality services in a timely manner. A highly skilled team of professionals is always ready to meet expectations of even the most demanding clients. Looking to achieve that goal, we develop and expand. Therefore we now offer not only DTP but also multimedia, eLearning and voiceover services.

Languages: All  
Hornet Design Studio  
Bydgoszcz, Poland  
+48525290553  
Email: office@hornetdesign.eu  
Web: http://hornetdesign.eu

**Education**

**44th Internationalization & Unicode Conference**

For over 29 years the Internationalization & Unicode® Conference (IUC) has been the preeminent event highlighting the latest innovations and best practices of global and multilingual software providers. This annual conference is where experts and industry leaders gather to map the future of internationalization, ignite new ideas and present the latest in technologies and best practices for creation, management and testing of global web, and multilingual software solutions. Recent conferences have provided specific advice on designing software for European countries, Latin America, China, India, Japan, Korea, the Middle East and emerging markets. Join us in Santa Clara, October 14-16, 2020, to contribute your ideas!

The Unicode Consortium  
Mountainview, CA USA  
Email: rick@unicode.org,  
Web: www.unicodeconference.org/ml-bg

Ad on page 17

Ad on page 47

Ad on page 21
**E4.**

**Total Solutions for Your Business**

E4NET is a total localization solutions provider including translation, DTP, recording, and specialized in Asian localization covering all major Asian and regional tier 3 languages. We have 20+ years of successful localization production experience with major projects for IBM, Microsoft, SAP, Oracle, HP, LG Electronics, Panasonic and more. E4NET is now providing patent translation services to the Korea Institute of Patent Information and translating life science projects including clinical protocols and reports. We are continuously developing and applying innovative technologies such as machine translation and associated customer services throughout our production process to maximize production/service efficiency. ISO 9001: 2015, ISO 27001 certified.

Languages: 60+

E4NET Co., Ltd. Seoul, South Korea
82-2-3465-8532
Email: info@e4net.net
Web: www.e4net.net

---

**Mobico – by Saltlux Inc.**

Mobico is the new brand name of Saltlux’s technical communication services, and is also the name of the predecessor company to Saltlux, established in 1979 as Korea’s first TC business. What started as a small enterprise concentrating on creating Korean manuals and East Asian language translations evolved into a one-stop service provider for all your needs in the world of business today, including multilingual translation, localization, DTP, TW and MTPE. The relentless pursuit of progress and perfection results in the use of state-of-the-art technology and processes, which in turn lead to superior translation quality with shorter turn-around times and therefore to greater customer satisfaction.

Languages: More than 70 languages

Saltlux, Inc. Seoul, South Korea
+82-2-2193-1725
Email: sales@mobico.com
Web: www.mobico.com/en

---

**ORCO S.A. Localization Services**

Established in 1983, ORCO celebrates this year its 35th anniversary. Over the years, ORCO has built a reputation for excellence and gained the trust of leading companies, such as Oracle, IBM and Carrier for the localization of their products. Our core business activities include technical, medical, legal, financial, marketing and other translations, software and multimedia localization, as well as localization consulting. We cover most European languages and our client list includes long-term collaborations with international corporations, government institutions, banks, private enterprises, NGOs and the European Union. ORCO is certified according to ISO 17100 and ISO 9001 quality standards.

Languages: Greek and European languages

ORCO S.A. Athens, Greece
+30-210-723-6001
Email: info@orco.gr
Web: www.orco.gr

---

**EuroGreek Translations Limited**

Established in 1986, EuroGreek Translations Limited is Europe’s leading Greek localizer, specializing in medical, technical, financial and legal translations from EN/DE/FR/ES to EL and EL to EN. Our aim is to provide high-quality, turnkey solutions, encompassing a whole range of client needs, from translation to localization, desktop publishing and testing. Our DTP department covers all Latin and Cyrillic alphabet-based languages, in addition to Greek, at very competitive rates. All of EuroGreek’s work is produced in-house by a team of 30 highly qualified specialists and is fully guaranteed for quality and on-time delivery.

Languages: Greek to/from English; French, German, Spanish to Greek

EuroGreek Translations Limited Athens, Greece
30-210-9628-559
Email: production@eurogreek.com
Web: www.eurogreek.com
Ad on page 71

---

**EuroGreek Translations Limited**

Established in 1983, ORCO celebrates this year its 35th anniversary. Over the years, ORCO has built a reputation for excellence and gained the trust of leading companies, such as Oracle, IBM and Carrier for the localization of their products. Our core business activities include technical, medical, legal, financial, marketing and other translations, software and multimedia localization, as well as localization consulting. We cover most European languages and our client list includes long-term collaborations with international corporations, government institutions, banks, private enterprises, NGOs and the European Union. ORCO is certified according to ISO 17100 and ISO 9001 quality standards.

Languages: Greek and European languages

ORCO S.A. Athens, Greece
+30-210-723-6001
Email: info@orco.gr
Web: www.orco.gr

---

**Enterprise Solutions**

**STAR Group**

Multiple Platforms

STAR is a leader in information management, localization, internationalization and globalization services and solutions such as GRIPS (Global Real Time Information Processing Solution), PRISMA (Smart Content Services), STAR CLM (Corporate Language Management) including Transit (Translation & Localization), TermStar/WebTerm (Terminology Management), STAR MT (Machine Translation), CLM WebEdit (Web-based Translation & Review) and MindReader (Authoring Assistance). With more than 50 offices in 30 countries and a global network of prequalified freelance translators. STAR provides a unique combination of information management tools and services required to manage all phases of the product information lifecycle.

Languages: All

STAR AG (STAR Group headquarters)
Ramsen, Switzerland; +41 52 742-9200
Email: info@star-group.net, Web: www.star-group.net

STAR Group America, LLC Lyndhurst, OH USA
216-691-7827, Email: lyndhurst@star-group.net
Ad on page 11

---

**Localization Services**

**CRESTEC**

Number 1 LSP that you should know

Crestec is a world-class language service provider. CSA Research ranks Crestec as a global top 5 LSP in the technology sector and global top 6 in the manufacturing sector. Headquartered in Japan with a network of over 18 sites in the US, Europe and Asia, we offer a one-stop comprehensive solution for global communications from marketing content creation and technical writing to localization, printing and studio/shooting production.

Languages: Japanese, Simplified Chinese, Traditional Chinese, Korean, French, German, Italian, Spanish, Dutch, Swedish

Crestec Long Beach, CA, USA
612-986-3108
Email: aki.ito@crestecusa.com
Europe: Amsterdam
+31 205854640
Email: sales@crestec.nl, Web: https://crestecusa.com

---

We add local flavour to your global business
**RWS Moravia**

RWS Moravia is a leading globalization solutions provider, enabling companies in the IT, consumer electronics, retail, media and entertainment, and travel and hospitality industries to enter global markets with high-quality multilingual products and services. RWS Moravia’s solutions include localization, testing, content creation, machine translation implementations, technology consulting and global digital marketing services. Our customers include eight of Fortune’s Top 20 Most Admired Companies, and all of the “Fab 5 Tech Stock” companies from 2017. Our global headquarters is in Brno, Czech Republic, and we have local offices in Europe, the United States, Japan, China and Latin America. To learn more, please visit us at www.rws.com/moravia.

Languages: over 250

RWS Moravia

USA HQ: Thousand Oaks, CA USA, +1-805-262-0055
Europe HQ: Brno, Czech Republic, +420-545-552-222
Email: moravia@rws.com
Web: www.rws.com/moravia
Ad on page 72

**Vistatec**

We have been helping some of the world’s most iconic brands to optimize their global commercial potential since 1997. Vistatec is one of the world’s most innovative, progressive and successful localization solutions providers. Headquartered in Dublin, Ireland, with offices in Mountain View, California, USA. Think Global.

Languages: All

Vistatec

Europe: Dublin, Ireland, 353-1-416-8000
North America: Mountain View, CA USA 408-898-2364
Email: info@vistatec.com, Web: www.vistatec.com
Ad on page 20

**Spark**

**A creative approach to localization**

At Spark, we fuse creativity with technology to provide a localization+ service. Helping brands like Microsoft, Disney and Netflix to develop and deliver consistent global branding and hyperlocal content creation from marketing activation to packaging. Providing creative expertise across the entire consumer journey, continuous workflow solutions and seamless integration into business systems: making a real difference to your localization needs.

Languages: 120+

Spark - Brighter Thinking

Europe: London, UK
4+44 (0)207 602 9119
hello@sparkbrighterthinking.com
www.sparkbrighterthinking.com
Ad on page 13

**Joint National Committee for Languages**

The joint National Committee for Languages and the National Council for Languages and International Studies (INCL-NCLIS) represent the interests of over 140 member organizations, associations and companies in virtually all aspects of the language enterprise — education PreK-20, research, training, assessment, translation, interpreting and localization — to the US government. The mission of INCL-NCLIS is to ensure that all Americans have the opportunity to learn English and at least one other language.

Joint National Committee for Languages – National Council for Languages and International Studies

Garrett Park, MD USA, 202-580-8684
Email: info@languagepolicy.org
Web: http://languagepolicy.org

**Translation Commons**

Translation Commons is a nonprofit US public charity powered by translators. We are a volunteer-based online community aiming to help our language community thrive and bridge all the sectors within our industry. We facilitate cross-functional collaboration among the diverse sectors and stakeholders within the language industry and instigate transparency, trust and free knowledge. Our mission is to offer free access to tools and all other available resources, to facilitate community-driven projects, to empower linguists and to share educational and language assets.

Translation Commons Las Vegas, NV USA (310) 405-4991
Email: krista@translationcommons.org
Web: www.translationcommons.org

**Translators without Borders**

Originally founded in 1993 in France as Traducteurs sans Frontieres by Lori Thicke and Ros Smith-Thomas to link the world’s translators to vetted NGOs that focus on health and education, Translators without Borders (TWB) is a US nonprofit organization that aims to close the language gaps that hinder critical humanitarian efforts worldwide. TWB recognizes that the effectiveness of any aid program depends on delivering information in the language of the affected population.

Languages: 190 language pairs

Translators without Borders CT USA
Email: info@translatorswithoutborders.org
Web: www.translatorswithoutborders.org
Ad on page 68

**Nonprofit Organizations**

**Translation Commons**

Las Vegas, NV USA
(310) 405-4991
Email: krista@translationcommons.org
Web: www.translationcommons.org
Ad on page 68

Keep an eye on the industry

Join more than 6,500 language and global communications professionals who receive the latest news in their inboxes every month.

multilingual.com/newsletter
Memsource

Memsource is a leading cloud-based translation management system that enables global companies, translation agencies and translators to collaborate in one secure, online location. Internationally recognized for providing an easy-to-use, yet powerful CAT tool combined with a TMS, Memsource processes two billion words per month from over 200,000 users around the world. Manage your translation projects in real-time in an intelligent platform that accepts over 50 file types and offers REST API out-of-the-box CMS connectors and powerful workflow automation to save time and money. Join localization professionals from around the world who rely on Memsource to streamline their translation process. To start your free 30-day trial, visit www.memsource.com.

Languages: All
Memsource Prague, Czech Republic
+420 221 490 441
Email: info@memsource.com
Web: www.memsource.com
Ad on page 43

Plunet BusinessManager

Plunet develops and markets the business and workflow management software Plunet BusinessManager — one of the world’s leading management solutions for the translation and localization industry. Plunet BusinessManager provides a high degree of automation and flexibility for professional language service providers and translation departments. Using a web-based platform, Plunet integrates translation software, financial accounting and quality management systems. Various functions and extensions of Plunet BusinessManager can be adapted to individual needs within a configurable system. Basic functions include quote, order and invoice management, comprehensive financial reports, flexible job and workflow management as well as deadline, document and customer relationship management.

Plunet GmbH Berlin, Germany
+49 (0)30 322 971 340
Email: info@plunet.com
Web: www.plunet.com
Ad on page 37

Smartcat

At Smartcat we believe the translation industry should be better for everyone. We connect linguists, companies and agencies to streamline the translation of any content into every language on demand. Our platform helps you build and manage translation teams, and puts your translation process on autopilot from content creation to payment. The unique features of Smartcat are our marketplace, where you can find translators for any language with one click; our CAT tool, translation using an AI-assisted platform, a team management with full control of your team, suppliers and content and payment automation; pay vendors easily across the globe. You can start experiencing the next generation of translation technologies and boost your translation business efficiency from day one.

Languages: All
Smartcat Cambridge, MA USA
Email: support@smartcat.ai
Web: www.smartcat.ai
Ad on page 23

Smartling

Smartling Translation Cloud is the leading translation management platform and language services provider to localize content across devices and platforms. Smartling’s data-driven approach and visual context capabilities uniquely positions brands for efficiency. Seamlessly connect your CMS, code repository, and marketing automation tools to Smartling’s TMS via prebuilt integrations, web proxy, or REST APIs. No matter the content type, Smartling automation tools help you do more with less. Smartling is the platform of choice for B2B and B2C brands, including InterContinental Hotels Group, GoPro, Shopify, Slack, and SurveyMonkey. The company is headquartered in New York, with offices in Dublin and London. For more information, please visit Smartling.com.

Smartling New York, NY USA
1-866-707-6278
Email: hi@smartling.com, Web: www.smartling.com
Ad on page 2

Consoltec

Multiple Platforms

Consoltec offers FlowFit-TMS, a web-based translation management system that helps you simplify and optimize your projects, while reducing your administrative costs. FlowFit can also be used for many other project types. FlowFit provides fully customizable web portals for clients, providers and project management. Get an accurate overview of your teams’ workload in real time and select the best available providers. Manage your clients, contacts and internal/external providers effectively with the new CRM features. Use Timesheet to track the time spent on projects and tasks. Connect seamlessly to your favorite CAT tools (memoQ, SDL Studio, LogiTerm) and get comprehensive reports that provide enhanced insight on production, productivity, costs and translation memory efficiency.

Consoltec Montreal, Québec, Canada
(+1) 514 312-2485
Email: info@consoltec.ca
Web: www.consoltec.ca

Localize

Localize offers a full-featured, cloud-based content and translation management system that features advanced translation workflows, allowing content managers and translators to propose, review, and publish translations with ease. For companies without in-house translators, we provide access to high-quality, on-demand translations through our network of professional translators. Our easy to install plugin fits neatly into your existing technology stack. The technology powering the Localize Platform was built from the ground up to minimize the need for engineers in the localization process. This reduces costs by enabling nontechnical personnel to manage the localization workflow. Getting started is easy. Start your free trial today!

Languages: All
Localize Kingston, NY, USA
(415) 651-7030
Email: sales@ localizejs.com
Web: https:// localizejs.com
Ad on page 51

Pragmatic Technologies

Multiple Platforms

Pragmatic Technologies offers FlowFit-TMS, a web-based translation management system that helps you simplify and optimize your projects, while reducing your administrative costs. FlowFit can also be used for many other project types. FlowFit provides fully customizable web portals for clients, providers and project management. Get an accurate overview of your teams’ workload in real time and select the best available providers. Manage your clients, contacts and internal/external providers effectively with the new CRM features. Use Timesheet to track the time spent on projects and tasks. Connect seamlessly to your favorite CAT tools (memoQ, SDL Studio, LogiTerm) and get comprehensive reports that provide enhanced insight on production, productivity, costs and translation memory efficiency.

Pragmatic Technologies Montreal, Québec, Canada
(+1) 514 312-2485
Email: info@consoltec.ca
Web: www.consoltec.ca

Plunet GmbH Berlin, Germany
+49 (0)30 322 971 340
Email: info@plunet.com
Web: www.plunet.com
Ad on page 37

Smartling

Smartling Translation Cloud is the leading translation management platform and language services provider to localize content across devices and platforms. Smartling’s data-driven approach and visual context capabilities uniquely positions brands for efficiency. Seamlessly connect your CMS, code repository, and marketing automation tools to Smartling’s TMS via prebuilt integrations, web proxy, or REST APIs. No matter the content type, Smartling automation tools help you do more with less. Smartling is the platform of choice for B2B and B2C brands, including InterContinental Hotels Group, GoPro, Shopify, Slack, and SurveyMonkey. The company is headquartered in New York, with offices in Dublin and London. For more information, please visit Smartling.com.

Smartling New York, NY USA
1-866-707-6278
Email: hi@smartling.com, Web: www.smartling.com
Ad on page 2

Ad on page 23
Translation Services into Italian

ASTW is an Italian language service provider that offers translation services into Italian. ASTW has historically gained extensive experience in patent translations and is now the provider of numerous international intellectual property consultancy firms. The other specializations include translations in the life science, legal and technical fields. ASTW also offers (light and full) post-editing services for pre-translated texts through machine translation in many areas. Other services include project and medical writing services in English and Italian.

ASTW Genova, Italy
+39 010 980 766
Email: info@astw.com
Web: www.a-stw.com/en/contact-us
Ad on page 32

iDISC Information Technologies

iDISC, established in 1987, is an ISO 9001 and ISO 17100 certified language and software company based in Barcelona with branches and teams in Mexico, Brazil, USA, Argentina, Bolivia and Guatemala. We have dedicated teams for web content, software localization and translation of technical, business, automotive, biomedical and marketing documents. Our software development engineers and translation teams provide high-quality and on-time production solutions that are cost-efficient, flexible and scalable.

Languages: Spanish (all variants), Portuguese (all variants), Catalan, Basque, Galician, Valencian, K’iche’, Quechua, Aymara, Guarani

iDISC Information Technologies, S.L. Barcelona, Spain
34-93-778-73-00
Email: info@idisctech.com, Web: www.idisctech.com

Wordbee Translator

Web-based

Wordbee is the leading choice for enterprises and language service providers that need to save money and make their company run more efficiently. Wordbee has the most complete feature set of any cloud solution: project management, portal, business analytics, reporting, invoicing and a user-friendly translation editor. Tasks such as project and workflow setup, job assignment, deadline calculation, multiple phase kick-offs and cost management can all be automated in the collaborative translation platform. Also, the Beebox connects CMSs, DMSs or any propriety database source with the TMS of the translation vendor or internal translation team.

Languages: All

Wordbee Soleuvre, Luxembourg
+352 2877 1204
Email: info@wordbee.com
Web: www.wordbee.com

birotranslations

Founded in 1992, birotranslations specializes in life science, legal, technical, IT and automotive translations into all East European languages (Albanian, Bosnian, Bulgarian, Croatian, Czech, Estonian, Hungarian, Latvian, Lithuanian, Macedonian, Polish, Romanian, Russian, Serbian, Slovak, Slovenian, Ukrainian). We have a long-term partnership with the world’s top 100 MLVs and many end-clients all around the globe. With our experienced project managers, extensive network of expert linguists and usage of the latest CAT tool technology, your projects will be delivered on time, within budget and with the highest standards of quality. For more information, please contact Mr. Matic Berginc (details below).

Languages: Eastern European languages

birotranslations Ljubljana, Slovenia
+386 590 43 557
Email: projects@birotranslations.com
Web: www.birotranslations.com
Ad on page 42

GlobalWay Co., Ltd.

As an industry-leading localization company in Korea, GlobalWay has been providing incomparable professional localization services with exceptional quality to partners all around the globe since 2003. We are here to offer language solutions including translation, voiceover, testing, DTP, and engineering services. Our highly qualified in-house linguists in each field of expertise, experienced engineers, and project managers will add value to your growing business. GlobalWay and its long-term global partners are ready to support you on the road to success. Are you looking for a reliable partner? Our doors are wide open for you. Should you need more information, please feel free to contact us.

Languages: 50+ more languages including Korean, English, Chinese, Japanese, German, Russian, Vietnamese, Thai, Indonesian

GlobalWay Co., Ltd. Seoul, South Korea
+82-2-3453-4924
Email: sales@globalway.co.kr, Web: www.globalway.co.kr

Educators:

Help further your students’ understanding of the intersection of language, technology and culture via articles written by experts around the world. Provide them with a digital subscription to MultiLingual with our compliments.

Contact us at subscriptions@multilingual.com to learn more about utilizing MultiLingual in your classroom.
MediLingua Medical Translations BV
Leiden, Netherlands
+31-71-5680862
Email: info@medilingua.com
Web: www.medilingua.com
Ad on page 41

MediLingua Medical Translations
MediLingua is one of the few medical translation specialists in Europe. We only do medical. We provide all European languages and the major languages of Asia and Africa, as well as translation-related services to manufacturers of devices, instruments, in vitro diagnostics and software; pharmaceutical and biotechnology companies; medical publishers; national and international medical organizations; and other customers in the medical sector. Projects include the translation of documentation for medical devices, surgical instruments, hospital equipment and medical software; medical information for patients, medical students and physicians; scientific articles; press releases; product launches; clinical trial documentation; medical news; and articles from medical journals.
Languages: 45, including all EU languages

Rheinschrift Language Services
Outstanding localization requires world-class experience. Rheinschrift gives your business a native voice in the German-speaking world. We offer more than 20 years’ experience providing translations and localizations for software and hardware manufacturers as well as for the sectors of business, technology, legal matters and medicine/medical applications. Our services also range from glossaries, post-editing, project management and desktop publishing services to many other related services. Rely on Rheinschrift to deliver the most competent translations and meet your deadline, whatever it takes.
Languages: German to/from major European languages

Translated.
Professional translation services made easy. Crafted by expert humans, powered by technology, efficiently delivered. We have delivered 1.2 million translations in 150 languages to 134,091 clients in 40 macro-domains since 1999, powering the globalization strategy of the most demanding clients. We work hard to make translation services more effective, by enhancing our production processes with great technologies and talented people. A perfect example is T-Rank®, the system that instantly matches your content with the most qualified translator for the job. We offer a wide range of linguistic services that cover all your future needs: Google Ads translation, software localization, subtitling, and APIs to integrate human translation. We open up language to everyone.
Languages: 150 languages and 40 areas of expertise.
Translated Rome, Italy
+390690254001
Email: info@translated.com
Web: www.translated.com
Ad on page 3

Translation Tools

memoQ
memoQ is a technology provider that has been delivering premium solutions to the translation industry since 2004. For almost 15 years, memoQ has been dedicated to delivering innovation through diverse developments that today help hundreds of thousands of freelance translators, translation companies and enterprises worldwide. Having simplicity and more effective translation processes in mind, memoQ combines ease of use, collaboration, interoperability and leveraging in one single tool. Discover a new world with memoQ, and let our team help optimize your translation processes and make your business more successful.
Languages: All
MemoQ Budapest, Hungary
+3618088313
Email: sales@memoq.com
Web: www.memoq.com

SDL plc
SDL is the global innovator in language translation technology, translation services and content management. Over the past 25 years we’ve helped companies deliver transformative business results by enabling powerful, nuanced digital experiences with customers around the world. SDL is the leading provider of translation software to the translation industry and SDL Trados Studio is recognized globally as the preferred computer-assisted translation tool of government, commercial enterprises, language service providers and freelance translators.
Languages: All
SDL plc Maidenhead, United Kingdom
+44-1628-417227
Email: info@sltrados.com
Web: www.sdl.com or www.sdltrasdos.com
Ad on page 5

ADVERTISER INDEX

ASTW 32
bioranslations 42
Creative Words 55
Endangered Alphabets Projects 41
EuroGreek Translations Limited 71
Game Global 10
HanseMUG, Inc. 59
Hornet Design Studio 47
JFA 41
Internationalization & Unicode Conf. 21
Localize 51
LocWorld 17
MediLingua Medical Translations BV 41
Memsresource 43
Middlebury Institute of Intl. Studies 16
MultiLingual Computing, Inc. 70
Plunet 37
RWS Moravia 72
SDL 5
Smartcat 23
Smartling 2
Spark 13
STAR Group 11
Translated 3
Translators without Borders 68
Vistatec 20
Words change lives.

The world’s great humanitarian and development challenges are hampered by the simple struggle to communicate.

We use language to make a difference.

Become a TWB sponsor or supporter at translatorswithoutborders.org/support-us
The need for multilingual individuals is growing at a rapid pace in our workforce, in the public sector and, most importantly, in our classrooms — be they virtual or in-person. We are living in a more interconnected and interdependent world — a world where knowing a second (or even third) language is critical to being successful in a 21st-century global economy.

According to a 2019 survey by ACTFL, nearly one in four US employers acknowledged losing or being unable to pursue a business opportunity over a singular lack of language skills.

Statistics like these indicate an alarming trend. As the world language skills gap in the US economy continues to grow, the number of qualified teachers continues to fall. The language education profession is facing incredible challenges when it comes to teacher recruitment and retention. As of 2018, 44 states and Washington, D.C. currently report shortages of qualified language teachers.

“In today’s global workforce, it is essential that all students have equitable access to language education in the classroom. A critical step in helping to make this vision a reality is to ensure that the US expands and sustains its pipeline of well-trained language educators,” says Bill Rivers, executive director of the Joint National Committee for Languages.

The state of language education in the United States and the clear need for employees capable of speaking languages other than English would dictate that more funding — and funding that specifically addresses teacher recruitment and retention — is needed immediately to resolve the issue.

Yet the Trump administration’s budget request for the 2021 fiscal year would not only fail to provide adequate funding for language education initiatives, it would eliminate federal funding for several major programs designed to strengthen and promote them, including:

Title II of the Every Student Succeeds Act (ESSA): ESSA was signed into law by President Obama in 2015 and is critical to ensuring our students and teachers have the resources and support to succeed. Title II provides grants to state and local educational agencies. The purpose of these grants is to increase achievement, improve the effectiveness of educators and expand access to effective educators, especially to low-income and minority students.

Title IV, Part A of ESSA: another key section of ESSA, Part A of Title IV provides grants to
state educational agencies to fund educational programs such as world language education, arts education and college and career counseling, among others. Well-rounded curricula, as funded by Title IV, encourage engaged and collaborative students who in turn perform well in school and life.

**Title II of the Higher Education Act (HEA):** HEA designates funding to support the Teacher Quality Partnership grant program for colleges of education. These grants improve teacher education programs, strengthen educator recruitment and provide training for prospective teachers. This funding is especially crucial as the US faces a shortage of qualified educators.

**Title VI/Fulbright-Hays:** Administered by the International and Foreign Language Education office, Title VI and Fulbright-Hays fund grant and fellowship programs that strengthen language instruction, area/international studies teaching and research, professional development and curriculum development at the K-12, postsecondary and graduate levels. If eliminated, traditionally underserved and undereducated students will end up further behind their peers.

These cuts will only make it easier for those considering language teaching to choose another career path, further weakening an already fractured teaching pipeline. Employers will find it increasingly more difficult to fill critical positions tomorrow if we are unable to recruit those educators tasked with teaching students these language skills today. Anyone who values a robust global economy and our ability to compete in it as a nation should strongly oppose these cuts to the federal budget.

The trends and statistics are undeniable — languages matter. They have cultural, economic, strategic and other implications for our nation. To deny this would be irresponsible. The cuts proposed by President Trump would leave our students and businesses unable to compete with their international counterparts and would leave us less safe as a nation. Instead of turning our backs on those who advance language education, we should be doing everything we can to support them. [M]

---

**Give The Gift Of Knowledge**

– for free!

If you find information in any issue of *MultiLingual* that you know would benefit a friend or colleague, let us send them a copy.

Simply email freecopy@multilingual.com with the person’s name, postal address and the date of the issue you’d like to share.

They will receive a complimentary copy of the magazine along with a note, letting them know you were thinking of them.
Building your success

One project at a time
AI is evolving by leaps and bounds, as seen in recent advancements in machine translation and language quality tools. And businesses are grappling with how humans and machines will work together in this new landscape.

My mission is to create best-fit combinations of artificial intelligence applications plus human processes that drive our clients’ content life cycles.

www.rws.com/moravia