Industry Watch

How to make money in the translation business

ROBERT DALE

Arria NLG plc
e-mail: Robert.Dale@arria.com

(*Received 11 January 2016*)

Abstract

Machine Translation research suffered a major blow in the 1960s, but it came back with
a vengeance. From a commercial point of view, it's now a mature technology that many
Internet users take for granted. We look at where we are now, and consider the scope for
new entrants into the market.

---

1 An anniversary year

2016 marks the fiftieth anniversary of an important event in the history of Machine
Translation (MT). In 1966, after two years of work, the group of seven scientists who
constituted the US National Science Foundation's Automatic Language Processing
Advisory Committee (ALPAC) handed down a 124-page report that was, well,
somewhat negative about the state of MT research and its prospects.[1] The ALPAC
report is widely credited with causing the US government to drastically reduce
funding in MT, and other countries to follow suit.

As it happens, 2016 also marks the tenth anniversary of the launch of the
Google Translate web-based translation service, which was soon followed in 2007 by
Microsoft's Translator. Google says its translation service is used more than a billion
times a day worldwide, by more than 500 million people a month. In mid-2015, one
market research report estimated that, by 2020, the global MT market will be worth
$10B.[2]

Not a bad turnaround in outlook, even if it did take a few decades.

---

[1] You can read the report here: `http://www.nap.edu/read/9547`. A detailed analysis by
John Hutchins is available at `http://www.hutchinsweb.me.uk/MTNI-14-1996.pdf`.
[2] See `http://www.pressreleaserocket.net/global-language-translation-software-market-is-expected-to-grow-10-6-billion-by-2020-acute-market-reports/264986`.

2 MT is special

In the portfolio of language technology applications that are the focus of interest of this journal's readership, MT occupies a special place. MT was the goal of one of the very first experiments in Natural Language Processing. In 1954, the Georgetown–IBM MT system automatically translated sixty Russian sentences into English, leading its authors to claim that within three or five years, MT might be a solved problem. You can still find the original press release on the web; it's a fascinating read, with its detailed description of a 'brain' that 'dashed off its English translations . . . at the breakneck speed of two and a half lines per second.'[3]

MT is also special because it's one of the first areas of Natural Language Processing where statistical methods took hold in a big way. Although the idea of statistical MT was first raised by Warren Weaver in a 1949 memorandum,[4] it was IBM's influential statistical MT work in the late 1980s and early 1990s that caused researchers to sit up and take notice. I think it's reasonable to claim that the perceived successes of Statistical Machine Translation (SMT) have been a major driver for the application of statistical techniques in other areas of Natural Language Processing since that time.

And MT is special because it's possibly the most accessible form of language technology in terms of the popular understanding. It can be a struggle to explain to the layperson exactly what text analytics is, or why it is that grammar checkers and speech recognisers make mistakes. But most people get what MT is about, and can see that it might be a hard thing to do; many people have struggled with learning a second language. Nobody doubts the value of a technology that can take one human language as input and provide another as output.

In fact, universal translators have been a staple of science fiction, and thus part of the popular imagination, since at least 1945.[5] Devices that can translate languages have played a role in many popular sci-fi TV shows. You can even guess someone's age bracket by the movie or TV show whose name comes to mind when you mention the idea—for me, it's *Star Trek*, where the back-story is that the Universal Translator was first used in the late twenty-second century for the translation of well-known Earth languages.

From where we stand now, *Star Trek*'s creator, Gene Rodenberry, looks to have been just a bit on the cautious side with his predictions. Perhaps he had read the ALPAC report: the Universal Translator first showed up in a 1967 episode of the show.

3 Where we are now

So, sixty-seven years after Warren Weaver first floated the idea of MT, we have instant translations of web pages on tap for more than fifty languages; Skype

---

[3] See `http://www-03.ibm.com/ibm/history/exhibits/701/701_translator.html`.
[4] See `http://www.mt-archive.info/Weaver-1949.pdf`.
[5] Murray Leinster's 1945 novella, 'First Contact', is often credited with the first appearance of a universal translator.

Translator translates voice calls in English, Spanish, French, German, Italian and Mandarin; and Google's Translate app translates foreign-language signs and menus using your phone's camera. We've come a long way, although it's sobering to consider just how long it has taken.

This is not to say that MT is a solved problem, at least in the sense of fully automatic high quality translation (FAHQT). It's widely recognised that current SMT systems are fine for 'gisting', but you probably wouldn't want to use them to translate a legal document.

But it has also been recognised from the early days that some form of human involvement in MT, either in pre-editing or post-editing, is necessary if you want to achieve high quality output. Bar-Hillel made this observation in the 1950s, and an entire translation industry has grown up around the use of translation memories and other supporting tools and machinery.

From the consumer's point of view, you might say the translation problem *has* been solved. If you want something quick, free and somewhere short of perfect, use one of the many web translators. Often that's all you need. If you want higher quality, there are plenty of services built on the use of sophisticated tools that reduce the cost of what would otherwise be a completely manual process.

From this vantage point, the translation market is now quite mature. The integration of MT into search engines like Google and Bing, where it clearly adds immense value, makes it hard for a new entrant to compete. The established presence of 'full service' translation providers like SDL and Systran likewise provides a challenging barrier to entry for anyone who wants to develop better tools for translators.

But you'd be wrong to assume this means that there's no life left in commercial MT innovation.

4 Human versus machine

Over the last year or so, a recurrent theme in blog postings and news stories is that 'the robots are coming'. We are frequently warned that more and more white collar occupations will disappear as software gets smarter.

Translators, though, would seem to be safe. The US Bureau of Labor Statistics predicts a forty-six per cent increase in translation job opportunities between 2012 and 2022. That's much higher than the eleven per cent average growth across all careers.[6]

It may not be what the US Bureau of Labor Statistics had in mind, but it looks like some proportion of that work is going to surface in a new 'gig economy', where an Uber-style workforce is available to translate your content on-demand. Crowd-sourcing means that the human translation expertise required can be found for any language pair at more or less any time of day or night.

A number of companies have sprung up to facilitate this marketplace. If you want to get some translation done, you are probably interested in some or all of

---

[6] `http://www.bls.gov/ooh/media-and-communication/interpreters-and-translators.htm`.

three things: low cost, fast turnaround and quality results. The third of those is a little harder to measure than the first two, so it's not surprising that the quantitative measures are the ones promoted most by the vendors in this space.

For example, Gengo (`http://gengo.com`) offer 'people-powered translation at scale', across thirty-four languages. They claim that ninety-five per cent of requests are started within 120 minutes and completed in an average of one hour, but a typical user will see their project begin within seven minutes and finish within thirty-seven minutes. The per-word translation cost varies from six cents to twelve cents, depending on the quality required.

Conyac (`https://conyac.cc/en`) offers a similar service, promising results in as little as ten minutes. One Hour Translation (`https://www.onehourtranslation.com`), whose website proudly states 'Human Translation Only', has a neat online calculator where you specify various parameters that characterise your requirements and you get a detailed time and cost estimate.

It's not always clear what technology these companies are leveraging, but it's a safe bet that they see their competitive advantage being the use of smart tools that improve the productivity of their armies of human editors. If you've got some ideas around supercharged UI for translation memory tools and the like, this might be where you want to position yourself. But you'll have to compete with the incumbents in terms of an installed base of on-demand translators: Conyac claim access to 50,000 translators, and Gengo and One Hour both claim 15,000; numbers which will no doubt be way out-of-date by the time you read this. As always, having a neat piece of technology is only one component of the solution.

The services just mentioned are built around leveraging the capabilities of translation tools, often making a virtue out of their reliance on human rather than MT. But there are also services that provide hybrid solutions that mix SMT and human post-editing.

Unbabel (`unbabel.com`) uses MT to do a first-cut translation, then assigns the results to a human translator to correct errors and fix stylistic inconsistencies. To speed up turnaround, Unbabel has developed Smartcheck, a tool that assists the translator by helping spot possible errors. Translate.com is another service that offers you the choice of MT (via Microsoft Translator) or professional human translations.

So, we're in an environment where SMT is now a freely-available resource. If you want to make money out of translation, you have to do it by complementing that basic technology with value-added services, generally focussed around SMT's acknowledged weak spot: quality. And you have to add that quality fast and cheaply.

5 Delivery models

The other aspect of MT technologies that has seen a fair bit of activity over the last year is concerned with how MT is delivered to the end user. The easier you can make it possible for someone to use your MT technology, the more likely you'll get customers.

KantanMT (`https://www.kantanmt.com`) provides a cloud-based platform that lets users build customised SMT engines. You combine your own training data with

KantanMT's Stock Engines, which consist of 'millions of words of highly cleansed training data'. Obviously this provides a degree of domain customisation that might mitigate some of the quality issues that come with using a generic SMT engine.

Data security and confidentiality is a deal-breaker for many SaaS solutions. For those who would rather keep their texts and the translation process in-house, Precision Translation Tools (`http://www.precisiontranslationtools.com`) provides Slate, a desktop port of the Moses SMT software that comes packaged with associated tools that make it easier to get up to speed, for both Windows and Linux.

Another approach is to make it really simple to use translation technology in conjunction with other content creation tools. For example, Lingotek (`http://www.lingotek.com`) provides a Translation Management System that integrates with Drupal and WordPress. Minimal Technologies offers WOVN (`https://wovn.io`), a simple website plug-in that focuses on easily-editable MT, with the scope to upgrade to human translation via Gengo, mentioned above.

Finally, a major selling point for translation solutions is to allow vendors to reach a larger audience than they can achieve using just their own language. A 2015 report from Accenture and AliResearch, Alibaba Group's research arm, projects that the global B2C cross-border e-commerce market will expand to \$1 trillion by 2020,[7] and that more than 900 million people around the world will be international online shoppers at that point. That's a lot of potential for MT, and not just for major commercial operations; there are a vast number of users of C2C shopping portals too. Hardly surprising, then, that eBay has been experimenting with MT since 2013, and acquired the MT arm of AppTek (`http://www.apptek.com`) in June 2015 as part of its cross-border strategy. Amazon subsequently acquired Safaba Translation Systems (`http://www.safaba.com`) in September 2015, presumably with similar uses in mind—although perhaps they're also thinking of MT further down the road in the context of Amazon Crossing (`https://translation.amazon.com/submissions`), which is already one of the largest publishers of the translated literature in the US.

6 Want a piece of the action?

As we've seen, MT is alive and kicking, but now is probably not a great time to be trying to introduce yet another SMT solution into the general marketplace. There's probably more scope for domain-targeted MT, where you stand a chance of producing better quality than the generic engines can provide. But as with any business, there are so many things you have to get right beyond just the technology, and many of the offerings identified above are worth exploring to see what works and what doesn't.

If I was going to invest in MT today, I'd be looking for a neat business idea that makes use of an existing MT API. There's already an extensive list of these at `http://www.programmableweb.com/category/translation`, so what are you waiting for?

---

[7] See `http://www.alizila.com/cross-border-e-commerce-to-reach-1-trillion-in-2020`.