Literary machine translation:

Are the computers coming for our jobs?

James Hadley

In the wake of the second world war, experiments in machine translation began. In 1947, American mathematician Warren Weaver laid out in a memorandum his vision for how digital computers might be used to translate human language. Earlier in the decade, a series of computers, including the Bombes and Colossus had been used by the Allies at Bletchley Park to decode Nazi messages. By comparing translation to decoding, it was not a great leap of imagination to conceive of machines being used to render messages from one language to another.

During the 1950s and early 1960s, research devoted to creating machine translation systems, primarily of the English-Russian language pair, was seen as a national security priority on both sides of the Iron Curtain. One of the most notable events of this decade was the 1954 Georgetown–IBM experiment, in which the automatic translation of more than sixty Russian sentences into English was carried out by a rules-based system. This experiment was seen as such a success at the time that it was confidently claimed that the problem of machine translation would be fully solved within three to five years.

Multiple exceptions to language rules

Rules-based systems, comprising bilingual dictionaries, along with logical rules for how to handle the textual information were based on the traditional language teaching methods. However, as anyone who has learnt a second language knows, language rules tend to come with multiple exceptions, meaning that these systems quickly became unwieldy, slow, and plagued with errors. In 1966, the Automatic Language Processing Advisory Committee concluded that despite significant investment, machine translation systems were not likely to reach the same standards as human translators in the near future, and that efforts should be moved towards developing tools to assist translators, what later came to be known as CAT tools, such as Trados.

Thus, for over a decade, machine translation research in the US slowed to a crawl. However, it continued in other countries, with the focus falling on a very small number of languages, such as English and French, as in the case of the METEO system used in Canada from 1977-2001 to translate weather forecasts between the country's two official languages. Around the same time, rules-based systems were being replaced by statistical machine translation (SMT) systems, which did not rely on manually coded rules, but on large corpora of parallel sentences which a computer uses to produce new translations. These systems originally worked word-by-word, and later came to work phrase-by-phrase. Thus, they work relatively well for high-resource, relatively similar languages pairs, where the huge number of parallel sentences required to build the corpora are available. However, in the case of languages with substantial differences in word order, or a limited availability of parallel data, they tend to fair less well.

Statistics versus neurons

By 2014, these were being superseded by systems using neural machine translation (NMT). These also rely on large corpora of parallel sentences in the two languages under consideration. However, neural systems are modelled on the way that neurons communicate in the human brain, where many small processes are brought together to create the final product. So, they depart from statistical systems in that statistical systems use their corpora as the ingredients for their translations, whereas neural systems use their corpora effectively to learn how to translate for themselves. These newer systems tend to produce translations more quickly and to a much higher quality – to the extent that if they are given enough training data, they can produce translations that are indistinguishable from texts produced by human translators.

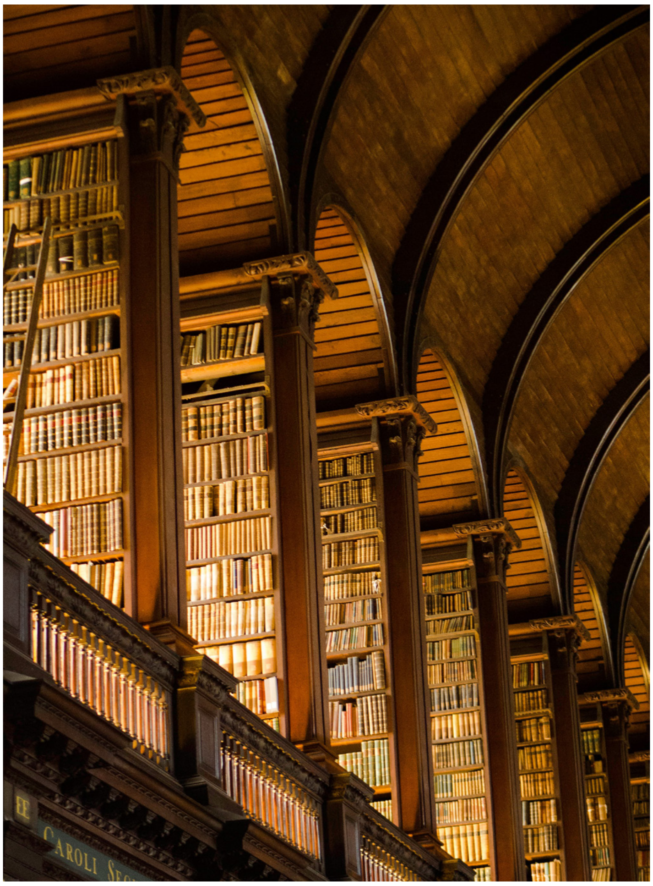
So, is it game over for the human translators? Well, no.

> ***"Authorial styles are not necessarily transferable"***

While neural systems are extremely effective at translating certain types of texts, especially those of a formulaic nature with short sentences, they are still very restricted in terms of what they can do well. This is because of technicalities that underlie the systems. In order to train a system, a large corpus of parallel sentences is required, and the system will perform better when it is trained on the kinds of sentence that it will be asked to translate. For example, a system trained on parallel sentences all taken from car manuals will likely perform very well at translating car manuals, and less well at translating cookery books. Getting past this issue is not as simple as just training a system on every kind of text, because the machine has no way of distinguishing which kinds of text it is trained on and dealing with in any given instance. Therefore, training a system on a wide range of text types will likely mean that the outputs for any of them will not be as strong as if the system were trained specifically on any one of them.

Millions of parallel sentences

This problem is not particularly grave when it comes to most technical texts, because the writing conventions that govern cookery books are not a million miles away from the conventions governing car manuals. So, while a system trained on all sorts of technical texts may not, statistically speaking, perform as well as a system trained on only one, the difference is often not so great as to cause serious issues. The same cannot be said for literature, however. In literature, not only are the writing conventions substantially different from many technical texts, these conventions differ substantially between authors, time periods, genres, and forms of literature.



'The Long Room' in the old library, Trinity College Dublin
Photo: Sarah Shaffe, Unsplash

Even though they are both kinds of poem, a sonnet is very different from a limerick. Even though they both fall into the fantasy genre of novels, *Harry Potter* is very different from *The Lord of the Rings*. The problem for machine translation systems is that for much literature, authorial styles are not necessarily transferable, and there is no precedent on which to build a system. Whereas there may be many parallel examples of contracts, for example, in the two languages with which a system can be trained, how is it possible to say what is the parallel of Dante in Swahili, or Tolstoy in Vietnamese? The nearest thing we could come up with would be a human translation of Dante into Swahili, or Tolstoy into Vietnamese. But a training corpus needs millions of parallel sentences to work effectively, which might work out as hundreds of books - many more than any one author is likely to have produced during their career. And practically speaking, if human translations for all these texts already exist, what is the point of training a system to translate the same texts in the same way again?

A corpus won't help with text style

It may be tempting to think that authorial style is not the end of the world. Surely, the "meaning" is what counts first? Well, that is not really the case when it comes to either literature or machine translation, where form and function are bound together. This was recently brought home to me in an experiment where we experimented with translating some poems from the *Arabian Nights* with a system which had been trained using the only parallel corpus available for Arabic-English, which is comprised mainly of Quranic translations and data from the UN. Although the vast majority of the words in the poems appeared in the training data, the style of the texts was so different from what the system was trained on that in most cases, it simply drew a blank.

A related point is that machine translation systems today work on the sentence-level, meaning that they translate one sentence in isolation, and then forget about it as soon as they move onto the next. Again, this is generally not a big issue when dealing with technical texts. But for literature, where ideas, metaphors, allusions and images can be recalled sentences, paragraphs, or even chapters later, the machines have a long way to go before they will be able to approach the skills of a human literary translator.

Software as the literary translator's assistant

For these and many other reasons, machine translation programmers are generally extremely tentative about what they expect from their systems and by when. Thus, what we are currently seeing is developers working on tools specifically to help literary translators. While some literary translators already make use of CAT tools, such as Memo-Q, many have not found these as relevant to them as to technical translators. But machines do have the ability to help with issues specifically relevant to literature.

For example, the QuantiQual Project is researching indirect literary translations produced by humans and machines. Indirect translations are translations of translations. For example, if a translation cannot be made directly from language A to language C, a mediating or "bridging" translation in language B might be used. While sentiments about whether this practice should take place at all have historically overwhelmed the fact that it has and does take place very widely, this project is interested in how the practice can be useful in helping us spreading knowledge and literature to languages that have historically been overlooked. One of the things the QuantiQual project is doing just now is to work out how the strengths of machine translation, in drawing on a very wide range of information sources; categorising technicalities; and identifying patterns, can be used to support human translators. The team is finding ways to help a translator who is faced with something like the poems in the *Arabian Nights*, and needing to render them into another language.

> ***"Any serious challenge to human literary translators is still a long way off"***

They are building a system which will not create translated poetry itself, but will give the human translator key details about the source text at a glance, which will allow them to work as efficiently as possible. For example, the software can tell the translator which of the

James Hadley *is Ussher Assistant Professor in Literary Translation at Trinity College Dublin, and the Director of the College's award-winning MPhil in Literary Translation master's degree. He is also the principle investigator on the QuantiQual Project, generously funded by the Irish Research Council's COALESCE scheme.*

James Hadley
Photo: Private Archive

rhyming patterns found in this type of poetry each text corresponds to, show where the rhymes, alliterations, and assonances are, tell the translator what the word counts, and average sentence lengths are, and give thesaurus-like glossaries in the target language for each of the words found in the poems. This way, the human translator is still the one choosing the most appropriate options and producing the translations, but the software is assisting them by allowing them to focus their attention on producing text, rather than on searching, and collecting supporting information from multiple sources. Compared to this highly complex text type, adapting a similar system to work with texts such as novels to assist translators in maintaining certain elements of style, such as sentence length, pronoun usage, or idiosyncratic word usages is a relatively small step.

So, while "never say never" is probably a good maxim, it is worth bearing in mind that pessimistic translators have been foretelling the arrival of their mechanical replacements since 1954. In translation studies and machine translation alike over the past seventy years, the more we have found out about translation, the more we have seen that it is much more complicated than we ever could have assumed. Any serious challenge to human literary translators is still a long way off, but we are already starting to see tools being developed that will assist literary translators in their work.