

The Proper Place of Men and Machines - Updated

Elliott Macklovitch

Université de Montréal, Montréal, Canada
firstname.familyname@umontreal.ca

Abstract. The central claim of Martin Kay’s famous article “The Proper Place of Men and Machines in Language Translation” is that one cannot automate what one does not fully understand – in this case, translation. In light of the impressive quality of the translations produced by current neural machine translation systems (henceforth, NMT), we return to that claim and examine whether it still obtains for NMT, which implements a variety of distributional semantics. Acknowledging the obvious, i.e. that machines clearly do not understand in the same way that humans do, we contend that NMT translations are indeed meaning-based. In support of that contention, we point to the success of certain NMT systems in achieving so-called zero-shot translation between languages where no explicit training data are available to them. We conclude by suggesting that these systems now require us to invert the proper roles of men and machines in language translation.

Keywords: neural machine translation, Martin Kay, language understanding.

1 Introduction

In 1980, Martin Kay, the late, great computational linguist, published his celebrated paper ‘The Proper Place of Men and Machines in Language Translation’.¹ Although the paper is famous today, it should be noted that it was not particularly well received at the time. In particular, it did little to slow or hamper the development of a host of machine translation (henceforth, MT) projects, all of which had as their aim fully automatic, high-quality MT. These included many commercial systems, such as Logos, Sys-tran, METAL and a host of Japanese projects, as well many university or otherwise publicly funded projects like EUROTRA. And while some may have claimed to serve merely as machine aids to human translators, in assuming the full responsibility for the translation process and relying on the human solely to clean up the machine output, they all fell within the purview of Kay’s critique.

‘The Proper Place’ is a remarkable paper for many reasons, not the least of which is the verve and sting of Kay’s prose. To cite just one example:

“There was a long period – for all I know, it is not yet over – in which the following comedy was acted out nightly in the bowels of an American government office with the aim of rendering foreign texts into English. Passages of innocent prose on which it was desired to effect this

¹ The paper first appeared as a Xerox research report in 1980 and was later republished in the *Machine Translation* journal vol. 12 (1997), as well as in the collection *Readings in Machine Translation*, MIT Press (2003). The page numbers of citations included here are those of the *Machine Translation* version.

delicate and complex operation were subjected to a process of vivisection at the hands of an uncomprehending electronic monster that transformed them into stammering streams of verbal wreckage. These were then placed into only slightly more gentle hands for repair.” (p. 5)

2 Kay’s Credo

In ‘The Proper Place’, Kay approvingly cites the previous work of Yehoshua Bar-Hillel, who, as far back as the early 1950’s, had convincingly argued for the non-feasibility of fully automatic, high-quality machine translation (coining in the process the well-known acronym FAHQT); see Bar-Hillel [1]. Bar-Hillel’s well-known thought experiment is based on the need to access and reason over unpredictable and potentially infinite amounts of real-world knowledge, even to translate a sentence as innocuous as “the box is in the pen”, a requirement he felt was and would remain absolutely unattainable for a translating machine. However, this did not prevent him from supporting machine translation as an excellent vehicle for fundamental linguistic research. What he deplored was the misguided promise that MT could replace human translators, or even serve to enhance their productivity in the short or medium term.

Kay fully agrees with this position, but the principal argument he advances against FAHQT in ‘The Proper Place’ comes from a different angle – that of a computer scientist.² Kay’s explanation for “the stammering streams of verbal wreckage” produced by the MT systems of his day is quite simple: “it happens when the attempt is made to mechanize the non-mechanical or something whose mechanistic substructure science has not yet revealed. In other words, it happens when we attempt to use computers to do something we do not fully understand. History provides no better example of the improper use of computers than machine translation.” (p.4) And later in the paper: “There is a great deal that computer scientists and linguists could contribute to the practical problem of producing translations, but, in their own interests, as well as those of their customers, they should never be asked to provide an engineering solution to a problem they only dimly understand.” (p.5)

What Kay is clearly implying in these statements is that our current understanding of translation, and perhaps more generally, our understanding of how human language works, is very partial and definitely insufficient to allow us to provide coded instructions to a machine on how to translate texts between two natural languages. For many years, in my own classes on translation technology, I used to refer to Kay’s argument in an effort to be explain to my students the poor quality of most machine translation output. Simplifying somewhat, what I would tell them was this: you can’t automate what you don’t understand. Or put another way: if you do attempt to automate what you don’t understand – in this case, translation – then this is the kind of output you obtain.

² In one of his many postings, Kay served as Chair of the Department of Computer Science at the University of California at Irvine.

3 That was then and this is...

2014 is the year that is generally cited as marking the advent of neural machine translation (NMT), the seminal articles often mentioned being Bahdenau et al. [2], Cho et al. [3], and Sutskever et al. [4]. So NMT has been with us for nearly ten years now, and it definitely has continued to improve over that period. Indeed, the best of today's NMT systems have reached a level that is so impressive that I have no qualms in asserting – and not just for the sake of the argument that follows – that FAHQT has finally been achieved. At first blush, this may sound like an overblown claim, and so we need to carefully consider what it does *not* entail. It does not mean that today's NMT systems always generate perfect translations; on the contrary, it is well known that they do occasionally produce translations that are incorrect in one way another, and sometimes bafflingly so. (But then again, what human translator can honestly claim to never making mistakes?) Nor is it to assert that the error-free translations they most often provide cannot be improved by a human revisor. (And again, the same is true for human translators.) Granting these provisos, it must be admitted that the output of today's best NMT systems³ bears absolutely no resemblance to the output provided by previous generations of machine translation, even the relatively recent statistical MT systems. NMT translations are almost always grammatically correct and idiomatic; and in most cases, they do convey the essential meaning of the source text being translated. That is precisely the goal that Bar-Hillel established for machine translation when he coined the term FAHQT.

4 Translation and Understanding

Here is another of Kay's bold quotations from 'The Proper Place':

"To translate is to re-express in a second language what has been understood by reading a text. Any purported solution to the problem that does not involve understanding in sense is, at best, ad hoc and therefore subject to the linguistic objections already alluded to." (p.7)

As argued in the previous section, I will assume that current NMT systems have indeed achieved the long-elusive goal of FAHQT. It then seems to me that one of two conclusions must necessarily follow: either such systems do indeed incorporate the kind of "understanding in sense" that Kay is alluding to in the above paragraph; or if they do not, then Kay's basic tenet is incorrect, since these systems continue to demonstrate that they can adequately translate what they do not understand.

Let me begin by stating the obvious: understanding is not a simple, monolithic notion that would allow one to unequivocally assert that, yes, NMT systems do understand the texts that they process; or no, they do not. Rather, understanding is a murky, loosely

³ In speaking of the *best* NMT systems, I am referring to those that have been trained on very large quantities of high-quality data. This is the case for English and French, the language pair that I work with. NMT output on language pairs for which the training data is insufficient will necessarily be less good.

defined concept that surely comprises different levels and admits of many different definitions. Take, for example, my own understanding of the artificial neural networks that underpin all NMT systems. For a translator and an old-school linguist⁴ like myself, these systems seem prodigiously complex and are exceedingly difficult to fully grasp. That said, it appears that the vectorized embeddings that play such a central role in NMT can be traced back to the linguistic theory known as distributional semantics, whose central postulate was famously encapsulated by JR Firth as “you shall know a word by the company it keeps”⁵; see Firth [5]. These word embeddings certainly manage to encode a great deal of information, semantic and otherwise, about each lexical unit in the vocabulary, including many (or all?) of the words it tends to cooccur with. From that point in the neural architecture, however, things become rather mysterious for me, as these lexical embeddings are then merged in hidden layers into sentence embeddings and ultimately converted into numerical representations that are projected into an abstract multi-dimensional space.

The overall encoder-decoder architecture of NMT systems is often said to be much simpler than that of the previous generation of statistical MT systems; nevertheless, the inner workings of NMT remain opaque to most users.⁵ What are we to make of these vectorized sentence embeddings produced by the encoder of a neural MT sentence? One possible approach that I personally find helpful refers back to the famous Vauquois triangle, which the French MT pioneer first proposed in 1968; see Vauquois [7].

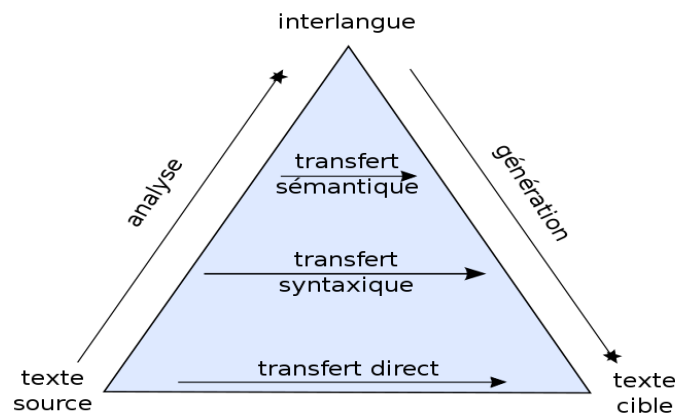


Fig. 1. The Vauquois triangle

⁵ An old-school *computational* linguist, moreover, who spent the greater of part of his professional career working on machine translation R&D projects.

⁵ Including some of the AI specialists who develop the systems! See Lee et al. [6] for their comments on the ‘hallucinatory’ output of their own NMT system.

In suggesting this schema, Vauquois' central point was this: the deeper the analysis that an MT system carries out on the source text, the less work its transfer component will have to do. As we progress up the triangle from first-generation direct MT systems, which conduct little or no analysis, to second-generation systems, which perform a syntactico-semantic analysis of the input, the number of necessary transfer operations is found to decrease.⁶ At the very tip of the triangle sits something conceived as a universal interlingua. Here, transfer disappears entirely; no transformations whatsoever are required to pass from one language to another. The classical examples often cited are Arabic numerals and chemical formulas: from NaCl, one can directly generate the linguistic expression 'sodium chloride' in English, 'chlorure de sodium' in French, and so on in other languages, without any further operations required.

Now one might be tempted to consider the vectorized embeddings produced by NMT systems as more elaborate instantiation of this interlingua. And indeed, there are certain AI researchers who make this claim quite explicitly; see, for example Lu et al. [8] and Escolano et al. [9]. The latter propose a multilingual MT systems that uses "multiple encoders and decoders for each language, sharing a *common intermediate representation*." The former describe "an attentional neural interlingua that receives language-specific encoder embeddings which are *agnostic to the source and target language*." (my emphasis in both cases) On the other hand, in Google's large-scale work on massively multilingual neural MT (see Arivazhagen et al. [10]), no mention is made of an interlingua. And when Angela Fan, the team leader of Meta's No Language Left Behind project, is directly asked the question about the status of their system's intermediate representations,⁷ she demurs, recognizing that the promise of multilingual MT has always been "some kind of multilingual space", but refuses to commit, saying instead that this is still an active area of investigation. (See her Youtube interview [11]; also [15] for a complete description of the NLLB project).

Notice, however, that the Vauquois triangle also includes a level just below the interlingua pinnacle, which is referred to as semantic transfer. Vauquois did not have much to say about this level of representation at the time, but in subsequent decades numerous formalisms have been advanced as candidates for semantic representations, although, as far as I know, few have been proposed within the framework of machine translation as providing the basis for deep linguistic transfer between languages.⁸ Here, I would like to (timidly) suggest that the vectorized numerical representations generated by neural machine translation systems are better viewed as constituting the kind of semantic representations that Vauquois had in mind, rather than as instantiations of a universal interlingua. That these vectorized embeddings are (largely) semantic in nature is only to be expected from an approach inspired by distributional semantics. Moreover, they have allowed for some relative success on zero-shot translation, i.e. the ability to

⁶ It is no accident, in other words, that Vauquois' schema is shaped in the form of a triangle and not as a rectangle, for example.

⁷ The question posed is incorrectly formulated in terms of Chomsky's universal grammar, but Angela Fan correctly interprets it to refer to a universal interlingua.

⁸ One exception that comes to mind is Lexical Functional Grammar, which has been used on a few machine translation projects. With how much success, I cannot say.

produce translations between language pairs for which no explicit training data is available. (Indeed, it is hard to imagine what else except for a semantic representation could possibly allow for translation between languages for which no explicit training data has been employed.) However, unlike the simple interlingual examples cited above, these zero-shot translations are often imperfect, meaning that further transfer operations (of some sort) would be required to transform them into target output that is fully meaning preserving. And that alone, in my view, is sufficient to disqualify them as being interlingual.⁹

Let us now return to Kay’s basic credo and once again ask the question: Does an NMT system understand the texts it is attempting to translate? Obviously, it does not achieve the same kind of understanding that human translators have when they translate a sentence. The system has no notion of the objects, processes and events that the words and sentences refer to in the real world, outside the texts.¹⁰ On the other hand, it does not seem to me unreasonable to claim that it does understand something very fundamental about translation. In proposing target sentence y as a translation of source sentence x , the system is implicitly making the claim that it understands the two sentences to mean the same thing. To take one simple example: an MT system needn’t comprehend what a ‘free-falling body’ refers to in the real world in order to know that the term is translated as ‘un corps en chute libre’ in French. Of course, it could be said that all machine translation systems have always been making this same implicit claim. True enough; but only NMT systems have managed to achieve a level of translation success that impels us to take this claim seriously.

We have been arguing that NMT systems do have a certain understanding of the texts they process and that the translations they produce are indeed meaning-based. That this understanding is not the same as that of a human translator is obvious enough,¹¹ but why should this matter? For years, we attempted to program the machine to emulate what we thought was the manner in which human translators operated, with very limited success. It was only when the rule-based, expert system approach was abandoned in favour of applying machine learning techniques to very large corpora of translated text that MT systems slowly began to improve.¹² And it was only when artificial neural networks were applied to that same task that machine translation output began to improve dramatically.

⁹ It is not sufficient, in other words, that these representations be “agnostic” between the source and target language; they also have to be adequate to directly generate a fully correct translation. For another take on the question, see do Carmo [12], his lecture on certain MT myths. While accepting the interlingual thesis, he is more skeptical of the claims made for zero-shot translation.

¹⁰ Piantadosi & Hill [13] convincingly argue that this does not prevent the representations learned by large language models from encoding important semantic information.

¹¹ Alan Melby is another who has argued that MT systems have no understanding of language, but merely “manipulate words mechanically”. See Melby and Kurz [14].

¹² The analogy that immediately comes to mind is human flight: it was only when people stopped flapping their arms like birds that human flight finally got off the ground.

That is the upside of the argument; but there is also a downside to the fact that NMT systems do not operate in any way similar to human translators. As mentioned above, the internal operations of these systems are excessively complex and difficult for those who use them to comprehend. For most working translators,¹³ an NMT system is very much a black box, one that we are still learning to work with. In previous generations of MT, translators enjoyed a modicum of control which allowed them, for example, to correct a dictionary entry and thereby alter the system's output. It is not yet clear how or to what extent users can modify the behavior of today's NMT systems, short of wholesale retraining. And retraining these data-hungry behemoths is no simple matter.

5 The place of the machine, reassigned

The second half of Kay's 'Proper Place' article is devoted to a concrete proposal for what he considers to be a more reasonable way of using computers to help working translators cope with the ever-increasing demand for their services. He calls his proposal a translator's amanuensis; on Pierre Isabelle's team at the CITI, we called a very similar project a translator's workstation.¹⁴ In both cases, what was being proposed was basically a multilingual word processor supplemented with a number of independent programs designed to assist the translator with various ancillary tasks, e.g., file format conversion, a personal glossary, spell checking, etc. Of course, all this sounds elementary today, but it has to be recalled that at the time of Kay's paper, the first popular personal computer (the IBM PC) had not yet been launched and very few translators had any experience working on a computer at all.

A key feature of Kay's proposal was its incremental nature:

"I want to advocate a view of the problem in which machines are gradually, almost imperceptibly, allowed to take over certain functions in the overall translation process. First they will take over functions not essentially related to translation. Then, little by little, they will approach translation itself. The keynote will be modesty. At each stage, we will do only what we know we can do reliably. Little steps for little feet!" (p.13)

Given Kay's negative assessment of machine translation, it is somewhat surprising to find that MT was not entirely banished from his amanuensis. But actually, the real target of Kay's attack is less machine translation itself than the manner in which these systems were employed at the time and the subordinate role that was left to the human translator. The standard *modus operandi* corresponded to what Kay colorfully described in the quotation given on the first page above: texts were first processed by the MT system (which invariably ran on a mainframe computer) and then passed on to a translator for correction. In Kay's amanuensis, on the other hand, it is the translator who firmly sits in the driver's seat; they are in complete control of the translation process

¹³ As opposed to the AI specialists who develop these systems; and even they appear to struggle with the systems' opacity. See Bau et al. [16] for the description of a study that aims to control the artificial neurons that determine a particular NMT output.

¹⁴ On the CITI's workstation project, see Macklovitch [17].

and may, if so desired, request a machine translation of a portion of text, which they can then accept, post-edit or decide to ignore.

As we mentioned above, MT system output did not substantially improve for many years (if not decades) following the publication of Kay's article. Hence, it was only natural that this secondary, optional recourse to machine translation which Kay ascribed to MT remain in effect, at least on those projects where translators had a say. Even today, within many translation environment tools that incorporate both translation memory (TM) and machine translation, priority is routinely given to the former over the latter. Concretely, this usually means that if a match for a given segment is found within the TM database, it is the one that is inserted by default into the system's editor, on the grounds that the translation memory contains human translations, which are assumed to be necessarily superior to those generated by an MT system.

Given the dramatic improvement in the quality of the translations generated by current NMT systems, I am not convinced that this division of labour between MT and TM remains valid today. At the very least, I believe that translators should always have access to the NMT output, alongside the TM output. Nor would I be surprised to learn that in a significant number of cases, translators choose to adopt or post-edit the MT output in preference to that retrieved from the memory.¹⁵

Yet, as previously mentioned, even the best NMT systems still occasionally produce erroneous translations, sometimes in the form of omitted content, less often in the form of wildly egregious (but grammatical) output. Because these errors remain by-and-large unpredictable, a qualified human translator will necessarily be required to revise all NMT translations that are either destined for wide dissemination or include content that could compromise security or potentially pose a danger.¹⁶ Why not just a proof-reader who is a native speaker of the target language? Because, paradoxically, it is much more difficult to detect the occasional semantic slip-up in the perfectly fluid output of NMT systems than it was to spot the often ungrammatical output that leapt off the page and demanded correction in the output of previous generations of MT. As the adoption of neural MT continues to grow, more and more translators will find themselves recruited to perform this kind of MT revision or, in the case of texts intended for publication, fine-grained MT post-editing. We may or may not like it, but for many of us, I am convinced that this is destined to become our proper place in the translation process in the coming years.

¹⁵ Particularly since the discrete segments stored in TM do not take the larger extra-sentential context into account, something that NMT systems are beginning to do. See for example Bawden et al. [18].

¹⁶ This point too was made by Kay in 'The Proper Place', where he argues against the often evoked statistical defense of MT, stating: "An algorithm that works most of the time is, in fact, of very little use unless there is an automatic way of deciding when it is and when it is not working." (p.10)

References

1. Bar-Hillel, Y.: The Present Status of Automatic Translation of Languages. *Advances in Computers* (1), 91–163 (1960).
2. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. <https://doi.org/10.48550/arXiv.1409.0473> (2014).
3. Cho, Kyunghyun, et al. "On the properties of neural machine translation: Encoder-decoder approaches." arXiv preprint arXiv:1409.1259 (2014)
4. Sutskever, I., Vinyals, O., and Le, Q. (2014). Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems (NIPS 2014)*.
5. Firth, J.R. (1957). A synopsis of linguistic theory 1930-1955. In *Studies in Linguistic Analysis*, pp. 1-32. Oxford: Philological Society. Reprinted in F.R. Palmer (ed.), *Selected Papers of J.R. Firth 1952-1959*, London: Longman (1968).
6. Lee, K., Firat, O., Agarwal, A., Fannjiang, C., Sussillo, D. Hallucinations in Neural Machine Translation. In: *Conference on Neural Information Processing Systems, Montréal, Canada. (2018)*
7. Vauquois, Bernard. A survey of formal grammars and algorithms for recognition and transformation in mechanical translation. In : *Ifip congress (2)*. 1968. p. 1114-1122.
8. Lu, Y., Keung, P., Ladhak, F., Bhardwaj, V., Zhang, S., Sun, J. A neural interlingua for multilingual machine translation. arXiv: 1804.08198 (2018).
9. Escolano, C., Costa-jussà, M., Fonollosa, J. Towards interlingual neural machine translation. arXiv: 1905.06831 (2019)
10. Arivazhagan, N., Bapna, A., Firat, O, Lepikhin, D., Johnson, M., Krikun, M., Chen, M., Cao, Y., Foster, G., Cherry, C., Macherey, W., Chen, Z., Wu, Y. Massively Multilingual Neural Machine Translation in the Wild: Findings and Challenges. arXiv:1907.05019v1 (2019).
11. Zeta Alpha Interview on Youtube: Angela Fan explains NLLB-200, <https://www.youtube.com/watch?v=IJZE7LikM3c> ; particularly around the 20-minute mark.
12. do Carmo, Félix. Debunking a few machine translation myths. CTS lecture uploaded on Youtube. <https://www.youtube.com/watch?v=Qap07fV5IgI> (2022).
13. Piantadosi, Steven T & Felix Hill. 2022. Meaning without reference in large language models. arXiv preprint arXiv:2208.02957
14. Melby, A. and Kurz, C. Data: Of Course! MT: Useful or Risky. *Translators: Here to Stay! MultiLingual* magazine (). <https://multilingual.com/issues/november-december-2021/data-of-course-mt-useful-or-risky-translators-here-to-stay/> (2021).
15. NLLB Team. No Language Left Behind: Scaling Human-Centered Machine Translation. <https://doi.org/10.48550/arXiv.2207.04672> (2022).
16. Bau, Anthony, et al. "Identifying and controlling important neurons in neural machine translation." arXiv preprint arXiv:1811.01157 (2018).
17. Macklovitch, E. The Translator's Workstation Project... in Plain Prose. In: *Proceedings of the 32nd Annual Conference of the American Translator's Association, Salt Lake City (1991)*.
18. Bawden, Rachel et al. Evaluating discourse phenomena in neural machine translation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics, Volume 1*, pages 1304-1313.