



HAL
open science

The disruptions of neural machine translation

Claire Larssonneur

► **To cite this version:**

| Claire Larssonneur. The disruptions of neural machine translation. 2019. <hal-03998505>

HAL Id: hal-03998505

<https://hal.science/hal-03998505v1>

Preprint submitted on 21 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

The disruptions of neural machine translation

Claire Larssonneur, University Paris 8

Abstract: According to the World Economic Forum, machine translation should outperform human translation by 2024... Quality translation based on machine learning has recently (2016-2017) been made available online and for free, via programmes like DeepL or Google Translate. Among the issues raised by these new technologies is the shifting geography and sociology of research, and the power plays behind: who has been developing these tools, who has been financing them? I have conducted an exhaustive survey of the 50 most relevant research articles on neural machine translation published in the year 2017 and available on Google Scholar. Not only does it reveal a geographical shift towards Asia but also a disciplinary shift from Humanities to Engineering. This redefines the imaginary of translation, from the scholarly figure of St Jerome to an AI assistant. It also impacts the epistemology of translation and more generally the attitude towards language, moving from a grammar-based approach to a neural/mathematical approach, as evidenced in the corpus. Such reframing of an essential mode of transmitting and exchanging meaning will have economic and political implications. Because neural machine translation requires powerful IT capacities and huge training corpuses, one can expect a further concentration of its major providers, leading to an oligopoly. And since machine translation is destined to be embedded in most apps and connected devices, issues of reliability, accountability and privacy are bound to surface.

Keywords: neural machine translation, research panorama, sociology geography, language, accountability, public utility

Claire Larssonneur is a Senior Lecturer at the University of Paris 8, France, specializing in Translation Studies and Digital Humanities. She acted as co-director of the international research programme “Le Sujet digital” (2012-2015), funded by the Labex Arts H2H, and guest-edited the “Digital subjectivities” issue of *Angles* (June 2018 <http://angles.saesfrance.org/index.php?id=954>). She also co-organized the panel “When Translation Goes Digital” for the IATIS Conference in July 2018 and is the founder of a research network on digital translation.

Claire.larssonneur@univ-paris8.fr

Introduction

According to the Future of Humanity Institute in Oxford, an estimate recently widely relayed by the World Economic Forum, machine translation should outperform human translation by 2024... and so translation and more generally the language services industry stand to be among the first to be disrupted by AI technologies.¹ The recent launch of free online neural machine translation tools, such as DeepL or Google Translate’s new version, because of their worldwide availability and enhanced performance, are bound to disrupt the translation market. These tools are already easily accessible either on the net or via apps, and the next step for the industry is

¹ Charlotte Edmond, “This is when a robot is going to take your job, according to Oxford University”, July 19, 2017. Posted on www.weforum.org. Available at <https://www.weforum.org/agenda/2017/07/how-long-before-a-robot-takes-your-job-here-s-when-ai-experts-think-it-will-happen/>

to embed them in many smart appliances like digital assistants or cars.² This obfuscates the whole production and mediation process, masking the materiality of translation. Yet the medium is essential, especially when dealing with such core social and political activity as language. In his study of digital economy, Olivier Bomsel pointed out that medium acts both as a materialization of the symbols it transmits, a tool to organize meaning, a tool to enact physical distribution and an exclusion tool upon which the definition of property and the attending rules rely.³ One might add that medium also acts as an identification tool (specifying the source, the author of the text) and as a venue for the conquest and assertion of power. As the materiality of translation disappears from view (and public debate), it becomes urgent to investigate the making of neural machine translation, to identify the genealogy and specificity of translation tools, to uncover the current sociology and geography of NMT agents, and examine its impact on our relation to language.

A Short History of Translation and its Imaginaries

To understand the disruptive force of neural machine translation, it is useful to retrace its genealogy within the history of translation. Translation, because it is a complex linguistic operation, was seen as a quintessentially human activity. It was usually represented by the figure of St Jerome, an elderly scholar surrounded by books, to which one may add the team of seventy-two translators who translated the Torah from Hebrew to Greek, around 270 BC. (insert picture 1 here) These representations showcased two important dimensions of the work: the length of time needed to translate (as indicated by Jerome's age) and the debates and discussions inherent to the process of translation (as indicated by the great number of books or contributors). Since the advent of computers and the digital processing of text, scientists have tried to devise automated machine translation tools. The first stage of machine translation at an industrial level⁴, starting in the late sixties and seventies, was based on linguistic rules, mostly lexical and morphological analysis coupled with the use of dictionaries. Its logic corresponded to that of the grammar book, a formal description aiming at some a posteriori rationalization of language. Rule-based machine translation processing was lengthy and error-ridden because it could not take into account the ambiguities and quirks of real-life language. Then toward the end of the eighties, researchers, notably from IBM and the German branch of Systran, developed statistical machine translation (SMT). This time the focus was on the production of native speakers, collected and ordained into huge databases. The whole process was based on matching databases of textual segments from different languages, without taking into consideration meaning and context. This shift to a mathematical approach in translation, embodied by the quest for the 'perfect match', may be argued to correspond to a larger trend within humanities, as identified by N. Katherine Hayles:

“The emphasis on databases in Digital Humanities projects shifts the emphasis from

² For instance the Google translate app for phones has been widely used by football fans in Russia during the 2018 World (Smith 2018) and a consortium of Japanese firms has developed Voice Tra, a free translating app targeting fans and tourists for the 2020 Summer Olympics.

³ Olivier Bomsel, *L'Economie immatérielle*, Paris, Gallimard, 2010, pp. 123-124 (my translation).

⁴ There are numerous accounts of the history of machine translation. The following is both precise and accessible to non-specialists : Ilya Pestov, “A History of Machine Translation from the Cold War to Deep Learning”, available at <https://medium.freecodecamp.org/a-history-of-machine-translation-from-the-cold-war-to-deep-learning-f1d335ce8b5> [Accessed August 20, 2018]

argumentation – a rhetorical form which historically has foregrounded context, crafted prose, logical relationships, and audience response – to data elements embedded in forms in which the structure and parameters embody significant implications.”⁵

But texts translated via SMT still contained many errors and seemed unnatural, again because the technology could not deal with homonyms and context. Then in a third phase, researchers integrated artificial intelligence technologies in the process, introducing machine learning to improve statistical machine translation by using predictive models. The reference to neurons, the very fact the algorithm gradually improves its performance via trial and error and the fluency of translations, all suggest the machine can work like a human brain, thereby legitimizing it and seemingly repatriating it within human-like activity. This obfuscates the mathematical nature of the process, its reliance on databases and computing power and its radical departure from natural language (NMT is based on the encoding and decoding of text into layers of semantic vectors that do not correspond to an existing natural language). One should add that the interfaces of DeepL and Google Translate are designed to give an immediate access to the translation result, simplifying the user experience into a mere sequence of copy/paste/click. In a sense those translation interfaces could be described as a form of artifice, theatrical tricks that overshadow the cumbersome process of translation. One could compare the artifice of neural machine translation to Marcello Vitali Rosati’s analysis of cloud computing:

“a metaphor that suggests that our data are immaterial, that they are nowhere, that they are light, and thus that they do not cost anything, that they are not on a particular hard-disk of a specific computer in a specific place, owned by a specific company. The ‘cloud’ metaphor as such is a way of forgetting all the economical, geopolitical, and social implications of a particular material infrastructure.”⁶

In this perspective, it could be fruitful to read the notion of artificial intelligence through the opposition between open and opaque processes of production, rather than the human versus machine paradigm. If one wants to understand the political and social implications of free ubiquitous machine translation, it is essential to investigate the hidden structures, interactions and actors of neural machine translation, the “set of dynamics that produce and structure [this specific] digital space”.⁷

How and Where is Neural Machine Translation Produced?

I have conducted on Google Scholar a systematic survey of the 50 most relevant research articles on “neural machine translation” published in 2017. The request yielded 2900 articles over a total of roughly 4800 articles on this topic: approximatively 60% of the output on Google Scholar was thus published in 2017. One striking feature of the sample was the glaring absence

⁵ N. Katherine Hayles, *How We Think: Digital Media and Contemporary Technogenesis*, Chicago, Chicago University Press, 2012, p. 39.

⁶ Marcello Vitali-Rosati, *On Editorialization, Space and Authority in the Digital Age*, Amsterdam, Institute of Network Cultures, 2018, p. 23.

⁷ This is an excerpt from Vitali-Rosati’s definition of editorialization (*Ibid.*, p.7): “Editorialization is the set of dynamics that produce and structure digital space. These dynamics can be understood as the interactions of individual and collective actions within a particular digital environment.”

of linguists or Humanities scholars: these research articles are written almost exclusively by researchers from Schools of Computing and Engineering, or departments for Computer Science.⁸ It entails a momentous shift of perspective on translation, moving from a focus on language as it is practiced daily and usually translated to content tailored to be processed swiftly by algorithms.⁹ Checking the institutional origin of those articles revealed a heavily polarized geography. On the one hand 46 per cent of the 50 most relevant articles on neural machine translation were written in English speaking countries, half in the USA (24 per cent) and half in other countries, mostly Ireland and the UK. On the other hand, 30 per cent were written in Asia, mostly China and Japan, with one coming from South Korea. Europe came third with 12 per cent, mostly from Germany and Poland. Closer examination of the data reveals another factor, the weight of corporate research. 40% of the articles in the sample were authored by researchers privately employed by such online tech giants as Google Brain, Facebook AI, Amazon or Microsoft, including Asian companies like Huawei Technologies, Tencent or Samsung Electronics. To which one should add a number of articles authored by academics but funded by those same corporations (10, one fifth of the sample). My hypothesis would be that the polarization of NMT research between the USA and China reflects the polarized landscape of tech giants, a finding consistent with the current American/Chinese contest for digital supremacy.¹⁰ This would also account for the relatively small number of language pairs under study: English is the one systematically recurring component, paired out either with Chinese or Japanese, or with a few European languages (German, Czech, Russian, Spanish, French). Such lack of linguistic diversity can be explained by technical constraints since neural machine translation requires training on large pre-existing corpuses of aligned segments, which are scarce. But it reinforces the hypercentrality of the English language¹¹ to an unprecedented level, and leaves out languages such as Portuguese or Arabic spoken by hundreds of millions of users¹².

The survey also revealed translation was seen as part of a wider attempt to improve the processing of multimodal content, especially when matching text, image and voice¹³ for the

⁸ When not Computing Science, the most common heading for those research centres would be Language Technologies, for instance at Carnegie Mellon or Dublin City University, but they are clearly labelled as computational linguistics and deliver MSCs, not MAs.

⁹ Artificial Intelligence technologies thus contribute to widen the gap between Translation Studies and Computational Linguistics that has been noted by Austermühl (2011) and Jimenez-Crespo (2013).

¹⁰ “FAANGS v BATS : the world’s tech giants are competing in new markets and the stakes are high”, *The Economist*, July 7th 2018, pp. 10-11. Also “A wall of money. Quite suddenly the tech firms have become the world’s biggest investors.”, Schumpeter column, *The Economist*, July 7th 2018, p. 52.

¹¹ Gisele Sapiro, “Center and periphery : Asymmetrical Flows of Translation”, in *A Companion to Translation Studies*, Chichester, Blackwell Wiley, 2014, pp. 82-94.

¹² Google Translate boasts it can translate over 100 languages, including Portuguese and Arabic but not as efficiently for all language pairs. DeepL focuses to this day on seven Western languages : French, English, German, Italian, Spanish, Dutch, Polish.

¹³ Within the sample, see Iacer Calixto, Qun Liu and Nick Campbell, « Incorporating Global Visual Features into Attention Based Neural Machine Translation” in *Computer Science - Computation and Language*, I.2.7, 2017 <https://arxiv.org/pdf/1701.06521.pdf> or the work of Elliot et al., funded by French and Dutch public institutions and Amazon Academic Research <https://arxiv.org/pdf/1710.07177.pdf>. For text-to-speech see Wang et al, “Tacotron: A Fully End-to-End Text-to-Speech Synthesis Model, Google Inc., in *Computer Science - Computation and Language* March 29, 2017 <https://arxiv.org/abs/1703.10135>

purpose of improving the performance of digital assistants like Alexa, voice-activated devices or crossmedia tagging on social media. Google Translate has clearly been updated for phone users, enabling finger writing or the translation of text embedded in pictures ([insert picture 2 here](#)) The sample under study has shown a profound disruption of the sociology and geography of translation. The polarization of research on NMT between the USA and China is cause for concern since those two countries have a history of extensive surveillance, censorship or fake news distribution, both through state bodies and private companies. This concern is bound to be amplified by the ubiquitousness of A.I. powered translation, freely available online and embedded as APIs in many appliances, websites and social media.

Economic, political and philosophical issues

The most glaring consequence of such disruption of the translation market is the increased power of a handful of tech corporations. Because neural machine translation requires powerful IT capacities and huge training corpuses,¹⁴ one can expect a further concentration of its major providers, leading to an oligopoly of global language service providers such as Systran, Star or Omnicore, together with the main tech giants such as Google, Amazon, Weibo etc. In this rarefied landscape, issues of reliability, accountability and privacy are bound to surface. One can cite the recent leakage scandal in Norway: Lise Randeberg revealed to the press in September 2017 she had found via Google search “notices of dismissal, plans of workforce reductions and outsourcing, passwords, code information and contracts” that had been fed by various companies to Translate.com and stored on the cloud.¹⁵ Information leakage is only one chip of the iceberg here. If one reads their terms and conditions,¹⁶ all content fed into free translating tools belongs to the corporations operating them. This raises issues of copyright in some cases and constitute another intrusive foray into our private lives, for the data thus collected will serve to refine profiling, marketing and surveillance. NMT stands thus as another addition to digital monitoring technologies, widening up their already considerable grasp on our interactions.

Finally, neural machine translation, together with other AI-based language processing tools, may impact our relation to language. From a technical point of view, there is a shift from natural languages, messy and constantly evolving, to streamlined linguistic content and data produced for machine processing, corresponding to a form of “controlled language”.¹⁷ From a user point of view, the diversity and complexity of translation is reduced to a single transfer operation. DeepL’s interface for instance ([insert picture 3 here](#)) automatically detects the language of the user via the IP address or settings of his device, defines the standard length of a message as 5000 signs and through the pairing up of two identical windows, strongly suggests there is only one translation for one source text although natural languages allow for a plurality of translated versions. In the case of free translation tools, not only do “devices, digital platforms, tools, networks, and protocols simultaneously provide the context of the content and act as the

¹⁴ Kirti Vashee, “Understanding the economics of machine translation”, [Translation Spaces](#) Vol. 2 p. 139. DOI: [10.1075/ts.2.07vas](https://doi.org/10.1075/ts.2.07vas) [Accessed August 20, 2018]

¹⁵ Florian Faes, “Translate.com exposes highly sensitive information in massive privacy breach”, <https://slator.com/technology/translate-com-exposes-highly-sensitive-information-massive-privacy-breach/> [Accessed August 20, 2018]

¹⁶ Google “Privacy and terms”. <https://policies.google.com/terms#toc-content>. Last updated: 25.10.17. [Accessed August 20, 2018]

¹⁷ Miguel Jimenez-Crespo, *Translation and Web Localization*, London and New York, Routledge, 2013, p. 199.

elements that structure this content.”¹⁸ they create the content. Indeed, from the point of view of communication theory, rather than messages being transmitted from human to human via a variety of media, we are now facing a situation where messages can originate either from machines or humans and be intended either for humans or machines (such as indexing robots or virtual assistants). In other words, the media becomes both sender and receiver. When an increasingly greater proportion of the texts and messages we are exposed to, both online or via smart appliances, are produced by machines, one can expect a massive standardization of languages. For instance, the more we rely on translation apps when abroad, as Dr. Joss Moorkens suggests, “the more we might be “trained” by those apps to speak in such a way to ensure the most accurate translation. “People will probably end up trained to speak in a restricted or unnatural way in order to achieve the best result,” he said.”¹⁹ Lack of diversity, standardization and control of language are iterative processes, so this trend may become massive. Finally, from a wider perspective, A.I. powered language production raises the issue of authority, not in terms of copyright but as the gateway to meaning. Vitali-Rosati reminds us that “authority does not guarantee that content – whether a sentence, an image, a video, web page, or any other fragment of information – corresponds to reality: authority is what creates reality.”²⁰

Conclusion

AI powered NMT is a two-faced tool: free ubiquitous quality translation may give an unprecedented boost to exchanges and prove extremely useful to many. But it entails a curtailing of the diversity of languages, and bypasses the element of debate and interpretation that was inherent to translation and that constituted its contribution to collective thought. The streamlining and standardization of language, the oligopolistic nature of the market, the lack of external control and accountability of those actors constitute threats not only to our privacy but also more generally to intellectual life. It might therefore be worth reflecting on creating Translation Commons or exploring ways to redefine translation as a public utility.²¹

¹⁸ Vitali-Rosati, *Ibid.*, p. 68.

¹⁹ In Rory Smith, “The Google Translate World Cup”, *New York Times*, July 13, 2018. Available at <https://www.nytimes.com/2018/07/13/sports/world-cup/google-translate-app.html>. [Accessed August 20, 2018]

²⁰ Vitali-Rosati, *Ibid.*, p. 83.

²¹ Vanessa Enriquez-Raido, “Translators as adaptive experts in a flat world: From Globalization 1.0 to Globalization 4.0?”, *IJOC: International Journal of Translation* Vol. 10, 2016, pp. 970-988.