Chapter 2

Human and machine translation

Dorothy Kenny

Dublin City University

This chapter introduces the reader to translation and machine translation. It attempts to dispel some myths about translation, and stresses the importance of translators in creating equivalence between source and target texts. Ultimately, the chapter aims to help readers construe human-produced translations as training data for machine translation. The chapter goes on to present some of the most useful distinctions made in machine translation: between types of machine translation systems and different uses of machine translation output. In particular, it attempts to explain contemporary machine translation as an application of the branch of artificial intelligence known as machine learning, and, more specifically, of deep learning.

1 What is translation?

This is a book about machine translation, which can be succinctly defined as translation performed by a computer program. This definition still leaves open the question, however, of what *translation* is. The reader should be made aware, at this point, that there is a vast amount of scholarship in the area known as *translation studies* that asks precisely this question, and that tracks the role of translation in diverse cultural, scientific and political arenas, to name just a few. It would be impossible to do justice to this rich field here, and the reader is referred instead to sources such as Baker & Saldanha (2020) for further information. We will content ourselves here by saying that most commentators would agree that translation is the production of a text in one language, the target language, on the basis of a text in another language, the source language. The notion of text is important. It refers to instances of real language use, whether spoken or written. In general, we expect texts to meet certain criteria: they should be coherent and

"hang together" properly; they should serve some kind of purpose, even if it is just to say "hello" to someone. We also usually have particular expectations regarding what texts will or should be like, given the particular language and context. This chapter, for example, hopefully meets the reader's expectations of a chapter in a collected English-language volume that is designed to be used as a textbook. It addresses a particular subject field or *domain*, namely machine translation, and adopts the conventions of a particular *genre*, that of a textbook.

The idea that translation involves texts is old hat to anyone who works in the area; it is so obvious that is doesn't need to be said. But in a world where most people don't think too much about translation, it is worth reminding ourselves that we translate texts and not languages. Languages are vast, complicated, abstract systems that are put to use in potentially infinite examples of human communication and expression. Texts are concrete instances of language in use. They normally have recognizable beginnings and endings, and even if individual languages seem to offer endless potential for creating sometimes unpredictable meanings and high levels of ambiguity, in any given text much of that potential simply falls away. It does not matter, for example, that *shower* in English can mean (1) a brief period of rain, (2) a device used for personal washing, or (3) a gift-giving party, all of which would be translated differently into a language like French, if what we are doing is translating a shower installation guide for a manufacturer of bathroom fittings. Unless the author is engaging in some witty wordplay, which is unlikely given the genre, we are dealing with the second meaning of *shower*. Focusing on texts rather than languages keeps things real, and manageable.

A second element of the definition of translation given above is the contention that translation involves the production of a text on the basis of another, pre-existing text. This clearly establishes translation as involving *a relationship between two texts*, commonly known as the *source text* and the *target text*.[1] Some commentators would go further than this and say that the relationship in question is one of having the "same meaning", but many philosophers and linguists – who understand meaning admittedly in quite sophisticated, technical ways – tend to shy away from claims of "same meaning" in translation. One reason for doing so is that it can be difficult to isolate the meaning of a text from the situations in which it is created and used. We might consider the meaning of a text

---

[1]A third element of our definition, of course, relates to the fact that source and target texts are in two different languages. We are thus concerned with *interlingual translation*. Some commentators, most notably Jakobson (1959), have recognized other types of translation, such as intralingual and intersemiotic, but a discussion of these categories is beyond the scope of this chapter.

to be what its writer or speaker wanted to say, but often we cannot be sure what they intended. Or we can associate meaning with our own interpretation of a text, but then we have to concede that other people might interpret the same text in a different way. A further issue that arises in the context of translation is that a perfectly valid target text may say more or less than its source text, simply because the language it is written in requires it to do so.

An example might help here. The opening line of a fairly recent memoir (Tammet 2006) is reproduced in example (1):

(1)   I was born on 31 January 1979 – a Wednesday.

Its translation into French (Tammet 2009) appears in example (2):

(2)   Je suis né le 31 janvier 1979. Un Mercredi.

Despite almost total word-for-word alignment between the two sentences, the French sentence actually says more than the English. It tells the reader that the writer, the *I* in English, is male, because if the writer was female, then the correct form in example (2) would be *née* and not *né*. Given certain tense forms, involving certain verbs, written French is obliged to signal the sex of the person in question.

But how does the translator into French know that the person saying "I" is male? This is, after all, the opening line of the book. Well, the book is a memoir, and the conventions of the genre require the enunciating subject to be the author of the memoir, and the translator knows whose book he is translating. It says it in the contract and on the cover of the book. The fact that French needs to specify the sex of a person in certain situations where this can be left vague in English does not cause the translator any headaches. It is a non-problem; but this very simple example shows two important things: the first – already mentioned – is that sometimes translations can mean more than their source texts. The second is that sometimes information that is required to translate a sentence cannot be found in that sentence. Rather one has to look into (1) the wider text – the front cover, for example – which is sometimes also called the *co-text*, the text that goes with a given fragment of text, or (2) the *context*, understood here as the wider situation that is relevant to the text, to find out how to proceed.

In other cases, a translation might say more than its source text not because the target language requires it, but because the genre does. In a study involving user interfaces for computer-aided design tools, Moorkens (2012) found that the single-word heading *Selecting* in English was commonly translated in a way that made explicit what was to be selected, yielding a variety of different translations, a sample of which is presented below, back-translated into English:

(3)   selection of polygon

(4)   selection of line

(5)   selection of ellipse

(6)   selection of rectangle

and so on.

This kind of explicitation, which results in translations saying more than their source texts, is not uncommon. The converse can also happen of course; in cases where it would be impossible or unusual for a target text to be as explicit as its source text, the translator can choose to leave out information. This can sometimes happen for language-typological reasons. For example, English belongs to a group of languages that frequently use verbs to describe the *manner* in which something or someone moves. Spanish, on the other hand, tends to use verbs to describe the *path* that is followed; it can encode the manner of motion in an adverbial phrase, but sometimes translators into Spanish will choose not to refer to manner of motion at all, as to do so would give it undue prominence, from the Spanish point of view. Slobin (2003) gives examples (7) and (8), by way of illustration. While the verb "stomped" in the English describes a way of walking in which the feet strike the ground heavily and noisily, the verb in Spanish "salió" simply captures the fact that the character in question has left the house.

(7)   He stomped from the trim house

(8)   Salió de la pulcra casa
      'exited from the trim house.'

There is a second way in which the Spanish sentence in (8) says less than its English counterpart in (7): the Spanish does not contain a subject pronoun equivalent to "he". This is because Spanish is predominantly a pro-drop language, meaning it can happily omit subject pronouns as most of the information they contain is available anyway from the ending on the verb in question, in this case, "salió", which indicates third-person singular, past tense. What's missing in Spanish but present in English is, of course, the gender of the subject. A reader of the Spanish text will, however, carry over knowledge of the (male) subject from the earlier co-text, and so they are not left in the dark. So by omitting the pronoun in Spanish, the translator has followed the norms of the target language and done no harm to the reader's ability to know what is going on in the novel.

The arguments and examples given above are intended to explain why so many scholars are reluctant to say that a source and target text have the same

meaning. What we are more likely to agree on is the idea that translations approximate their source texts. For all sorts of reasons, translators have to make decisions about what to prioritize when translating, what they need to say and what they should leave up to readers to work out for themselves.[2] The meanings that they help target-text readers to construct for themselves are likely to be compatible to a very large extent with the meanings that source-text readers construct, but in many cases they will not be identical. And that is generally not a problem.

But if we cannot call the relationship between a source text and a target text – or more probably snippets of such texts – one of "same meaning", then what can we call it? One answer is to call this relationship one of *equivalence.* Equivalence as a term has a chequered history in translation studies, but if it is understood as a relationship that emerges from the decision-making of a translator, a relationship that arises between two text snippets because the translator has deemed them to be of equal value in their respective co-texts and contexts, then equivalence can be a perfectly serviceable term. It allows us to say things like "salió" in example (7) is equivalent to "he stomped" in example (8). This equivalence is clearly not fixed for all eternity, and it certainly cannot be generalized to all other contexts in which the word "stomped" might appear, but this does not matter, if we concede that "salió" was a fair exchange for "he stomped" in this particular case.[3]

2 What makes translation difficult?

Books about machine translation frequently start with discussions of why translation is difficult, homing in on the kind of monolingual ambiguity and the systematic differences between languages alluded to above. Inter-linguistic differences might also be exemplified using cases where two languages are said to distribute meaning differently across the words in equivalent sentences. In examples (9) and (10), taken from the proceedings of the European Parliament, for example, the fact of liking something is expressed in the verb "like" in English, and what is liked is expressed in a complement to that verb "working with you". In German, what is liked is expressed in the verb "kooperiere", while the fact of liking is expressed in an adverb "gern".

---

[2]The examples we have given here are primarily caused by mismatches between linguistic systems, but a translator might chose to omit or change a detail in a source text for cultural reasons, to avoid confusing readers with unfamiliar references, or even to avoid offending readers or a censor. Or they may be constrained by space, as often happens in the production of subtitles.

[3]The idea of equivalence being based on exchange value is developed in the work of the translation scholar Anthony Pym. See, for example, Pym (2010).

(9)   I like working with you.

(10) Ich kooperiere gern mit Ihnen.
       'I cooperate happily with you.'

While such examples tell us something interesting about how psychological states are expressed in English and German, they don't really constitute a translation problem – for a human being, at least. A speaker of German with basic competence in English will be able to translate sentence (9) with little difficulty. Non-isomorphism between languages, the idea that languages are structured differently, does not in itself cause problems for translators.

Another linguistic phenomenon that is said to be tricky involves discontinuous dependencies, where two words that belong together are separated by one or more intervening words. "Send" and "back" in example (11), for instance, should be understood as a single lexical item meaning 'return'. Again, readers with a basic grasp of English generally have no problem working this out.

(11) Send your certificate of motor insurance back.[4]

Another frequently posited difficulty in translation is presented by idioms. Idioms are understood here as phrases whose meaning cannot be inferred on the basis of their constituent parts. Idioms, in other words, are non-compositional. A good example is "old hat", as used in 1 of this chapter. If you describe something as "old hat" you mean that it is so familiar that it has become tedious even to speak of it. The expression has nothing to do with head wear. Idioms, like other types of figurative language, where a word or expression should not be interpreted literally, can sometimes cause confusion for readers who have not encountered them before, but even if you did not understand "old hat" at first glance in this chapter, you are likely to have reasoned that the discussion had not moved on to millinery, and that a non-literal interpretation was in order. When this happens to a translator, she is likely to simply look the idiom up in an online dictionary, on the well-founded assumption that the expression is common enough to be included in such a dictionary. In other words, although the expression is non-compositional and figurative, it is still conventional. Despite the fact that the translator has hit a problem, in the sense that her flow has been broken, the problem is easily resolved and finding the solution will probably bring the translator considerable pleasure. (Translators, like all linguists, generally like learning new things about their working languages.)

---

[4]There are five intervening words in this case, making it one of the longest instances of a discontinuous phrasal verb in the sample of English known as the British National Corpus.

But even experienced translators will sometimes admit to difficulty in translating texts that are highly technical and for which they lack sufficient training. Whereas a translator with an educational or professional background in legal studies or practice might relish working on the translation of legislation, a specialist in automotive engineering will run (or drive) a mile from such work. It can also happen that, even within their own domain, translators can come across source texts that are badly written, or incomplete, or written in a way that makes them extremely difficult to understand. Or they may have no problem understanding the original, but face serious challenges in tracking down suitable terminology in the target language to label specialized concepts encountered in the source text. An unreasonable deadline, or a malfunctioning software program, are other factors likely to cause professional translators headaches. But you rarely hear a professional translator complain about linguistic ambiguity, non-isomorphism, discontinuous dependencies or non-compositionality.

The reason these phenomena appear so frequently in discussions of machine translation is, of course, that – in certain circumstances – they can cause problems for machines.

3 How do translators normally solve translation problems?

The above discussion mentioned a few real-world problems professional translators sometimes face. Here we look at a sample of problems that are related to the words on the page or, more probably, the screen. When a professional translator does not understand something in the source text, or cannot recall a specialized term in the target language, or is struggling to come up with a way of formulating an idea in the target language, she will usually divert her attention from the text at hand, and do some research. A translator grappling with the niceties of wastewater treatment, for example, may go to the website of various local authorities to see how they explain the technology involved. She might access one of the many publicly available termbanks to find an equivalent for a given term. She might consult other documentation produced by her client's company or speak to engineers at the company. She could consult with her colleagues, if she has any, or post a query to a translator's forum. The main thing is that most professional translators will realise when they have a gap in their knowledge, or need inspiration, and they will conduct conscientious research to address that gap, solve the translation problem and move on.

You might ask why it is so important to sketch how human translators work in a book about machine translation. The answer is twofold: firstly, in a very real way (elaborated upon by Rossi & Carré 2022 [this volume]) human translation sets the standard by which machine translation is judged, and anything that contributes to the maintenance of high quality in human translation is ultimately of relevance to machine translation. Likewise, human translation processes can help to put into sharp relief occasional deficits in machine translation. Human translation has a role to play, in other words, in both the evaluation of machine translation output and in the diagnosis of problems in that output. Secondly, and even more crucially, most contemporary machine translation relies on translations completed by humans to learn how to translate in the first place. This point is expanded upon below.

Before we close off our discussion of how human translators work however, we need to introduce a technology that has become indispensable for many translators: translation memory.

4 Translation memory

In the 1990s translators working in the burgeoning software localization industry found themselves translating texts that were either extremely repetitive in themselves or that repeated verbatim whole sections of earlier versions of a document. This was the case, for example, with software manuals that had to be updated any time there was a new release of the software. Rather than translate each sentence from scratch, as if it had never been translated before, they invented a tool that would store previous translations in a so-called *translation memory*, so that they could be reused. The tool, known as a *translation memory tool*, would take in a new source text, divide it into segments – sentences or other sensible units like headings or cells in tables – and then compare each of these segments with the source-language segments already stored in memory. If an exact match or a very similar segment was found, then the corresponding target-language segment would be offered to the translator for re-use, with or without editing. As translators worked their way through a new translation assignment, they would get hits from the translation memory, accept, reject or edit the existing translation and update the memory as they went along, adding their own translations for the source-language segments for which no matches existed. Over time, the translation memories grew extremely large. Some companies who were early adopters of the technology built up translation memories containing hundreds of thousands and then millions of *translation units*, that is source-language segments aligned with their target-language segments. Example (12) shows a simple

translation unit based on a headline (in English and German) taken from a translation memory consisting of data from the website of the European Parliament. It is presented in a format known as tmx (for "translation memory exchange"). The tags `<tu>` and `</tu>` open and close the translation unit, the tags `<tuv>` and `</tuv>` open and close each *variant* within the translation unit,[5] and the tags `<seg>` and `</seg>` open and close the *segment* or text string in that language.

(12)
```
<tu>
<tuv xml:lang="EN">
<seg>A common blacklist for unsafe airlines</seg>
</tuv>
<tuv xml:lang="DE">
<seg>Unsichere Luftfahrtunternehmen kommen auf eine schwarze
Liste</seg>
</tuv>
</tu>
```

Private translation enterprises also accumulated large translation memories, which came to be regarded as valuable linguistic assets that could help control translation costs and enhance competitiveness. International organizations such as the Institutions of the European Union adopted the technology and built up huge multilingual translation memories, which they in turn made freely available to computer scientists in the knowledge that they could support research agendas in natural language processing.

While translation memory was originally conceived as a way of improving, among other things, the productivity of human translators, it also eventually supported efforts to increase automation in the translation industry: on the one hand, translation memory tools enabled translation data to be created in great quantities and in a format that could be easily used in machine translation development (see below); on the other hand, the tools used to manage them provided an editing environment in which machine translation outputs could later be presented to human translators for editing alongside human translations retrieved from conventional translation memory.

Translation memories can be seen as a special type of *parallel corpus*, that is a collection of source texts aligned at sentence level with their target texts. In cases where translations were created without the use of a translation memory tool, translated texts could still be aligned with their source texts after the fact. So, for example, the translated proceedings of the multilingual European Parliament

---

[5]The first variant in this case is in English ("EN"), and the second in German ("DE").

were extracted from the web and aligned with each other to create the multilingual Europarl Corpus (Koehn 2005), which in turn gave a significant boost to machine translation research. Aligned parallel corpora do not have to be in tmx format. Often they take the form of files with thousands (or even millions) of lines, each line occupied by a single sentence, whose position in the file matches exactly that of its translation in another file in a given target language, so line x in the target language file contains the translation of line x in the source language file.

5 What is machine translation?

Based on the definitions given at the start of this chapter, we can say that machine translation involves the automatic production of a target-language text on the basis of a source-language text. As with other types of translation, we can expect the target text to allow an interpretation that is in most ways compatible with that of the source text. Although if we have already conceded that human translations can result in slightly different meanings to their source texts, then maybe we should allow machine translations to do the same. The important thing is that obvious divergences between source and target text, for example, where Japanese gives more information than English, should be motivated by the language pair, the genre or some other reasonable cause.

Machine translation was one of the first non-numerical applications of the digital computers that emerged in the aftermath of the Second World War. Early efforts to automate translation seem primitive by today's standards, although it has to be acknowledged that the protagonists were working with extremely limited resources in the 1950s and 1960s (Hutchins 2000). Nevertheless, automatic translation systems were in operation primarily in defence, government and international organisations by the late 1960s and 1970s, and by the end of the century their use was expanding in commercial settings. The technology became available to millions of internet users in 1997, when the American search engine AltaVista starting giving access to free, online machine translation under the Babel Fish name. In the decades since then, the internet has expanded rapidly, and now boasts some 4.66 billion users (Johnson 2021). By 2016, perhaps the best-known free, online machine translation system, Google Translate, was reported to have over half a billion users, translating over 100 billion words per day and supporting 103 languages (Turovsky 2016).[6] In combination with search

---

[6]As of May 2022, Google Translate supports 133 different languages, although to varying degrees (Caswell 2022).

engines like Google Search or Microsoft Bing, for example, machine translation can be used to expand a search and then to translate relevant foreign-language web pages back into the user's language.

But it's not all about web pages. Machine translation is also used in combination with technologies like automatic speech recognition and speech synthesis, or optical character recognition and digital image processing, allowing users to have spoken conversations in two or more languages, or read road signs written in unfamiliar writing systems, often using an app installed on their mobile phones. In some cases, these apps now even work offline and users can justifiably claim to be carrying a machine translation system in their pocket. Machine translation is also increasingly used in areas previously considered beyond the capacity of the technology, for example in audio-visual translation, to translate the subtitles of foreign-language movies and TV series into the language of a new market. Indeed, subscription video streaming services thrive on a model that brings the so-called long tail of lesser-known titles to a new audience, and many of these titles are lesser-known partly because they were originally made in a foreign language. Audio-visual content is thus becoming just the latest in a long line of commercial products whose markets can be expanded through machine translation. In the seventy or so years since its inception, machine translation has thus moved from being the preserve of governments and international organizations to being a mass consumer good.

Despite the undoubted usefulness of machine translation in the kind of scenarios addressed above and its capacity to do good in other, for example, humanitarian settings (Nurminen & Koponen 2020), it comes with some health warnings. First, just like human translators, machine translation systems can make mistakes. Errors might range from the amusing but trivial to the extremely serious (for example in healthcare, news translation or international diplomacy). Whole branches of research are thus devoted to *estimating* the quality that given machine translation systems are likely to produce, *evaluating* particular outputs, designing ways to correct errors by *post-editing* machine translation output or helping the machine produce better output in the first place, usually by *pre-editing* source texts to make them easier to translate. These areas are discussed in detail in Chapters 3 to 5 of this book. Machine translation also raises a surprising number of moral and legal issues, as addressed by Moorkens (2022 [this volume]) on ethics, and to a lesser extent by Carré et al. (2022 [this volume]) on machine translation for language learners.

Many casual users of machine translation may feel that they do not need to know much about any of these areas to get what they need from the technology: if you are simply using machine translation to get the gist of a text, to understand

the basic contents of a web page, for example, then this might be true. Such uses, which often fall under the heading of machine translation for *assimilation*, generally involve low-stakes, private use of the translated text in question, with little risk of reputational or other damage. If, however, you want to use machine translation for *dissemination*, for example to publish your blog in a second language, or to advertise your business, then it is wise to understand the risks involved and even to take measures to mitigate them. The ability to do so is a component of what is now known as *machine translation literacy* (Bowker & Ciro 2019). Other components include having a basic understanding of how machine translation actually works, and of the wider societal, economic and environmental implications of its use. While this might seem like esoteric knowledge, it turns out to be highly transferable, as contemporary machine translation is based on the same principles as a whole host of other technologies that are contributing to profound changes in many aspects of contemporary life, and especially how we work. In short, machine translation is now, for the most part, an application of machine learning, and more specifically of deep learning. These concepts are explained briefly below, and treated in greater depth by Pérez-Ortiz et al. (2022 [this volume]) on how neural machine translation works. If you are a translation student, a professional translator, or are employed in some other capacity in the translation industry, then you are probably strongly motivated to learn about what happens "under the hood" in machine translation systems. You are probably also interested in how you can get the best out of the technology, by customizing it for your needs. This is addressed in Ramírez-Sánchez (2022 [this volume]). The following paragraphs, on the other hand, should be read by anyone who is curious about how machine translation can be said to be the linguist's entrée into the wonderful world of machine learning.

6 Artificial intelligence, machine learning and machine translation

Contemporary machine translation is frequently mentioned alongside a number of other related concepts, including artificial intelligence, machine learning, artificial neural networks and deep learning, some of which can be difficult to differentiate for the uninitiated. Sources like Goodfellow et al. 2016 use a Venn diagram to explain how they relate to each other. Artificial intelligence (AI) is the most general category, represented by the biggest circle. It is often defined as the branch of computer science that aims to create machines – or more specifically computer programs – that can solve problems of the kind that would normally

require human intelligence. The machines in question don't necessarily have to *think* like humans, rather they need to *act* like an intelligent human would. They might be designed to solve fairly narrowly defined problems, like recognizing faces. Such goals are the stuff of *narrow AI*, also known, somewhat unkindly, as *weak AI*. So-called *strong AI* is a more aspirational undertaking. It would involve either *general AI* – in which machines would have human-like intelligence, be self-aware, able to learn and plan for the future – or *superintelligence*, which would involve intelligence that exceeds the abilities of any human. It is fair to say that translation, as practised by professional, human translators, requires the kind of intelligence that strong AI aspires to, but that such intelligence still remains beyond the capacity of machine translation systems.

6.1 Rule-based machine translation

One way to tackle the challenges of AI is to attempt to give a computer program all the knowledge it would need to solve a particular problem, and rules that specify how it can manipulate this knowledge. In the case of machine translation, for example, you can give the program a list of all the words in each of the source and the target languages, along with rules on how they can combine to create well-formed structures. You can then specify how the words and structures of one language can map onto the words and structures of the other language, and give the machine some step-by-step instructions (an *algorithm*) on how to use all this information to create translated sentences. This approach, known as *rule-based machine translation* (RBMT), dominated machine translation up until the early part of this century. When free online machine translation first became available in 1997, for example, it was based on RBMT (Joscelyne 1998). RBMT was beset by a number of problems, however. It was very expensive to develop, requiring highly skilled linguists to write the rules for each language pair and, like other knowledge-based approaches to AI (Goodfellow et al. 2016), it suffered from knowledge bottlenecks: it was simply impossible in many cases to anticipate all the knowledge necessary to make RBMT systems work as desired. This applies both to knowledge about language and knowledge about the wider world, so-called *real-world knowledge*.[7]

---

[7]Although RBMT has fallen out of favour generally, at the time of writing, it is still used in a small number of systems, especially for translation between very closely-related languages. See, for example, Apertium (Forcada et al. 2011).

6.2 Data-driven machine translation

This is where machine learning comes in. Machine learning is based on the premise that rather than telling a machine – or, more precisely, a computer program – everything it needs to know from the outset, it is better to let the machine acquire its own knowledge. The machine does so by observing how the problem it is intended to solve has been solved in the past. We have already seen how translation problems and their solutions can be captured at segment level in the translation units stored in translation memories and other parallel corpora. These translation units constitute the *training data* from which contemporary machine translation systems learn. This is why such systems are usually categorized as *data-driven.* And learning from data is what distinguishes machine learning from other types of AI.

Data-driven machine translation is divided into two types: *statistical machine translation* and *neural machine translation*, each of which is addressed below.

6.3 Statistical Machine Translation

Statistical Machine Translation (SMT) systems basically build two types of statistical models based on the training data:[8] the first model, known as the *translation model*, is a bilingual one in which words and so-called *phrases* found in the source-language side of the training data appear in a table alongside their translations as identified in the target-language side of the training data, and each source-target pairing is given a probability score. The ensuing structure is known as a *phrase table*. Table 1 contains an example of an excerpt from such a phrase table.[9]

Table 1: Excerpt from a Phrase Table showing *a me piace*, observed translations in Europarl and their probabilities

	English	Probability
a me piace	I like	0.78
a me piace	I should like to	0.11
a me piace	I admire	0.11

---

[8]A statistical model is a mathematical representation of observed data.

[9]The example is greatly simplified, as it shows only sensible Italian-English pairings. In reality, an SMT system would learn a translation model that contains lots of nonsensical pairings, most of which would, however, be assigned very low probabilities. It would also reserve some probability mass for previously unseen pairings.

The term "phrase" is something of a misnomer here however, as the strings in question don't necessarily correspond to phrases as commonly understood in linguistics. Rather they are *n*-grams, that is, strings of one, two, three or *n* words that appear contiguously in the training data. In the previous sentence, "appear contiguously" is a bigram, for example, and "appear contiguously in" is a trigram.

The second model, known as the *language model* is a monolingual model (or combination of models) of the target language. Again, it is based on *n*-grams. A trigram target language model, for example, would give the probability of seeing a particular word in the target language, given that you had already seen the two words in front of it. A trigram model could tell you the probability of seeing the word "gorgonzola" if you have already seen "I like" in the Europarl corpus, for example. It turns out to be 0.024, which means that while "I like gorgonzola" does occur in the training data (it actually occurs four times) there are many words other than "gorgonzola" that are much more likely to follow "I like".[10]

In SMT systems, the translation model is supposed to capture knowledge about how individual words and *n*-grams are likely to be translated into the target language, while the language model tells you what is likely to occur in the target language in the first place. What is really important from the current perspective, is that linguists don't have to hand-craft these models. Rather they are learned directly from the data by the machine in a *training* phase. In a second phase, called *tuning*, system developers work out the weight that should be assigned to each model to get the best output. Once the system is trained and tuned, it is ready to translate previously unseen source sentences. Translation (as opposed to training) is called *decoding* in SMT. It generally involves generating many thousands of hypothetical translations for the input sentence, and calculating which one is the most probable, given the particular source sentence, the models the system has learned, and the weights assigned to them.

SMT was state-of-the-art in machine translation for at least a decade up to 2015. It represented a huge advance compared to the RBMT systems that preceded it, but suffered from a number of deficiencies, most of them due to the fact that relatively short *n*-grams were used to build models and that *n*-grams in the same sentence were translated almost as if they were independent of each other. SMT performed particularly poorly on agglutinative and highly inflected languages. Other problems included word drop, where a system simply failed to translate a word, and inconsistency, where the same source-language word was translated two different ways, sometimes in the same sentence. By 2015, SMT was already

---

[10]The version of Europarl used here and in Table 1 is accessible through the Sketch Engine interface at sketchengine.eu.

being displaced by a competing approach to data-driven machine translation, the above-mentioned neural approach. Within a matter of two years the transition to neural machine translation was complete.

But if SMT is becoming obsolete, you might wonder why it is mentioned here at all. SMT is introduced here for the purpose of opening up the area of machine learning to the reader. SMT showed that machine translation systems that learned from data worked better than those that didn't. It thus paved the way for machine learning approaches in machine translation. SMT developers also made remarkable contributions to machine translation research, by promoting new methods and sharing their programs, but also by collecting translation data, from bilingual and multilingual parliaments, international organizations, the world wide web and so on, and sharing these data with the global research community. It should also be noted that SMT is still used in the translation industry, albeit in limited contexts: a supplier of machine translation services might, for example, first create an SMT system to see how viable the project is and whether or not it is worthwhile investing time and effort in subsequently developing a neural system.

Our main interest in discussing SMT is, however, to show that there is more than one way of learning from data[11] and, more importantly for our purposes, of representing those data. As we have seen, SMT represents translation knowledge in phrase tables, and target language knowledge in separate *n*-gram models. In such models, words (and strings of words) are still recognizable as themselves but, crucially, they are related to each other using probability scores. And it is these scores that allow the systems to work. The probability of any given target sentence being the translation of a given source sentence can be computed simply by multiplying the translation probabilities of its component *n*-grams as found in the phrase table, and the probability of any given target-language sentence occurring can be computed by multiplying the probabilities of its component *n*-grams, as indicated by the language model. A single equation can then be used to bring the different models together to compute the most likely translation.[12]

Another reason to address SMT is that doing so gives us a convenient excuse for introducing concepts such as *n*-grams, which turn out be to extremely important in other areas in natural language processing in general, and in machine translation evaluation in particular, as addressed by Rossi & Carré (2022 [this volume]).

---

[11]In fact, we have said very little about the specific algorithms used by SMT systems to learn. The interested reader is referred to Koehn (2010).

[12]The equation in question is based on Bayes Theorem, and SMT offers the translation scholar an entrée into the machine learning approach known as Bayesian optimization.

6.4 Neural Machine Translation

SMT had its heyday between 2004 and 2014. Most major users and suppliers of machine translation, including Google Translate (from 2007) and the European Commission (from 2010) were using the technology and in so-called *shared task evaluations*[13] SMT constantly came up trumps. Until 2015, that is. That year a neural machine translation (NMT) system developed at Stanford University beat a number of SMT systems – by a wide margin and on what was considered a difficult language pair, namely English-German (Bentivogli et al. 2016). The Stanford success heralded the beginning of what Bentivogli et al. (2016) call "the new NMT era." The excitement was palpable among researchers and especially in the press. Grand claims were made about the new technology, for example, that it was as good as professional, human translation and had thus reached *human parity*.[14] It was also claimed, with some justification, that NMT could learn "idiomatic expressions and metaphors", and "rather than do a literal translation, find the cultural equivalent in another language" (Marking 2016).[15] But while there is some truth in such claims, they should not be over-interpreted. An NMT system might indeed produce an idiomatic translation, but this is generally because the data it has learned from contain hundreds or maybe thousands of examples of that very translation. An NMT system (in this case Google Translate) does not *know* it is being idiomatic, or using a cultural equivalent, when it correctly translates the German idiom:

(13)   Ich habe die Nase voll.
        'I have the nose full.'

as

(14)   I'm sick of it.

Rather it is outputting what it has learned from data.[16]

But why is NMT so much better that SMT, if it is simply learning from data? Is that not what SMT was already doing? The answer lies in the kind of representations that NMT systems use and in the kind of models they learn.

---

[13]In shared task evaluations, computer scientists pit their systems against each other to see which performs best for a given language-pair and with different types of training data.

[14]This claim was famously made by researchers at Microsoft who had been working on Chinese-to-English translation (Marking 2016). It was contested by many commentators, including Toral et al. (2018).

[15]These comments were made by Alan Packer of Facebook in 2016 (Marking 2016).

[16]This particular translation has also been verified by Google Translate's user community.

6.4.1 Models in NMT

Let's start with models. A computer *model* is an abstract, mathematical representation of some real-life event, system or phenomenon. One use of such a model is to predict an answer to a previously unseen problem. A computational model of translation, for example, should be able to predict a target-language sentence given a previously unseen source-language sentence.[17]

We have already seen that SMT systems use probabilistic models of translation and the target language that are encapsulated in phrase tables and *n*-gram probabilities. NMT systems, in contrast, use models that are inspired, even if only loosely, by the human brain. They use artificial neural networks, in which thousands of individual units, or *artificial neurons*, are linked to thousands of other artificial neurons (let's just call them *neurons* from now on). In such a network, each neuron is activated depending on the stimuli received from other neurons, and the strength or *weight* of the connections between neurons. As Forcada (2017) explains, the activation states of individual neurons do not make much sense by themselves. It is, instead, the activation states of large sets of connected neurons that can be understood as representing individual words and their relationships with other words. The trick in training an NMT system is to learn precisely those weights that will result in the best performing model of translation, that is, the model whose activation states allow it to predict the best translations.

So how is this done? Like in all machine learning, the system learns from data. A neural model of translation is built step by step by exposing a learning algorithm to vast quantities of parallel data. In successive passes, the algorithm learns weights and keeps adjusting those weights, so that the predictions of the model it builds get closer and closer to a desired "correct" answer. More precise details of how this feat is accomplished are given by Pérez-Ortiz et al. (2022 [this volume]), and readers looking for a comprehensive technical discussion of NMT can also refer to Koehn (2020). It suffices to say here that data-driven machine translation is typical of machine learning in that it involves technologies that are developed to solve problems to which humans already know the answer and to which, in fact, humans have already supplied at least one, if not several correct answers. Such correct answers may be present in the training data or they may be arrived at through generalization from the training data. When a machine translation system is tested to see whether it is improving during training or to compare it to another system once training has finished, we also test by giving it

---

[17]When speaking of mathematical models, it is common to speak of *predicting* answers. For our purposes, however, there is little practical difference between *predicting* and *outputting* an answer, which is what machine translation systems do when they are actually being used.

a problem to which we already know the answer. Typically, we ask it to predict the translation of several sentences it has never seen before but for which we already have good (human) translations that we set aside specifically for this purpose.

When an NMT system has been trained to our satisfaction, it can be put into use in a real translation scenario. We no longer talk about "testing" the system, and instead talk about "using" it. When an NMT system is in actual use, most people say that the system is "translating". As with SMT, computer scientists also use the term *decoding* for the moment when an NMT system produces an output in the target language.

6.4.2 Representing words in NMT

We have already said that a mathematical model is a representation of some system, event or phenomenon. There is much debate over the status of mathematical and other scientific models, but that need not detain us here. We will take the view that a model represents something quite complex, with many interconnected parts, and that if we want to talk about a simpler or more granular entity – a number like 5, for example, or an object like an apple – we can simply use the generic term *representation* to refer to ways in which that entity is depicted.

Representations are important, because, as Goodfellow et al. (2016) point out, how ideas are represented affects what we can do with them, in computer science and in daily life. A good example of how representation affects human beings' performance is given by the difference between Arabic and Roman numerals. Most people would find it much easier to divide 125 by 5, for example, than to divide CXXV by V, even though CXXV and 125 (and V and 5) represent exactly the same quantity.

Words can also be seen as representing ideas. So the word *apple*, for example, might be understood as representing a particular type of fruit. Another way of representing the same fruit would be to draw a picture of it. The word and the picture would have different properties, of course, which allow you to do different things with them. You can spellcheck a (written) word, for example, but not a drawing.

In NMT yet another type of representation is used: the vector, which is a fixed-sized list of numbers. The word *apple* could be represented by a vector like [1.20, 2.80, 6.10] for example. To many people this seems incredible. It is difficult to see how a list of numbers can represent a word.[18] Things start to make slightly

---

[18]Note how we have shifted here from talking about the representation of ideas to the representations of words. What we have in training corpora are millions of identifiable words. They are what we try to represent in NMT.

more sense if we say that vectors are quite good at representing relationships between words. The vector [1.20, 2.80, 5.50], for example, could be the vector for *pear*. It differs from the vector for *apple* in just the last number. If we see the numbers in the vector as representing dimensions in an imaginary three dimensional space, this would make the words *apple* and *pear* very close to each other. And presumably they would both be far from less related words, like *helicopter* or *very*. Vectors have other interesting properties that make them particularly attractive to computer scientists. You can add a vector to another vector, for example, or multiply them and so on. Try doing that with the words themselves, or with drawings of apples and pears!

So how did our vectors for *apple* and *pear* end up so suspiciously similar in the above example? The truth is, we just made them up. In a real NMT scenario, we would get a computer program to *learn* suitable vectors for all instances of all words in our corpus directly from that corpus. (Remember, in machine learning, the computer program has to work these things out for itself, with or without human supervision.) The vector-based representations of words that the machine learns are called *word embeddings*. The reason why embeddings for related words end up looking similar to each other is that they are built up on the basis of where particular words are found in the training data. If it turns out that two words tend to keep turning up in the same or similar co-texts – both *apple* and *pear* occur very regularly before the word *tree* for example; both appear regularly after *peel*, *slice* and *dice* – then they will end up with similar embeddings.

Word embeddings are not built in one go, but rather in successive *layers*, as described in Pérez-Ortiz et al. (2022 [this volume]). An artificial neural network that has multiple layers sandwiched between its external layers is known as a *deep neural network*.

Deep learning, in turn, is simply the branch of machine learning that uses multiple layers to build representations. In a deep neural network, the external layers correspond to inputs and outputs of the network and are visible to the human analyst. The intermediary, or *hidden*, layers have traditionally been less open to scrutiny, however, giving deep learning a reputation for opacity, and encouraging some commentators to misleadingly use the word "magic" to describe the internal workings of deep neural networks. The mystique of NMT is added to when big tech companies report on their successes in building multilingual translation models, sometimes involving hundreds of languages, and which can cope with translation between languages for which there was no "direct" bilingual training data.[19] Researchers in AI have not been oblivious to problems caused by

---

[19]See https://ai.googleblog.com/2016/11/zero-shot-translation-with-googles.html and https://about.fb.com/news/2020/10/first-multilingual-machine-translation-model.

perceived opacity, however, and in the areas known as *explainable AI* (XAI) and *interpretable AI*, efforts are now being made to open up the "black box" of deep learning, so that its inner workings can be more easily understood by users, explanations can be provided for particular outputs and systems can be improved (see, for example, Vashishth et al. (2019)).

7 The advantages and disadvantages of neural machine translation

NMT is generally considered the best performing type of machine translation invented so far. It performs better than SMT, for example, because it can build up very rich representations of words as they appear in a given source text, taking the full source sentence into account, rather than mere *n*-grams. When it produces translations, an NMT system considers both these rich representations and the emerging target sentence at the same time. Because NMT handles full sentences, it is better at dealing with tricky linguistic features like discontinuous dependencies and it handles all sorts of agreement phenomena better than SMT.

But while contemporary NMT systems certainly handle full sentences, until recently, they did not look beyond the current sentence. This meant that they could not use information from a previous sentence to work out what a pronoun like "it" refers to in the current sentence, or that the understood subject of a Spanish verb is feminine (in cases of pro-drop in Spanish). This restriction to sentence-level processing can cause lots of other problems that only become apparent when users translate full texts rather than isolated sentences. The problem is currently being tackled by researchers working in the area of *document-level machine translation*, however (see, for example, Bao et al. (2021)). NMT can also output words that don't actually exist in the target language. Far more seriously, NMT output can be fluent but inaccurate. And when a translation looks and sounds good, one might neglect to check that it is compatible with the source text. Like other technologies trained on large quantities of existing text, it can also amplify biases encountered in the training data. A well documented form of the amplification of bias is the way in which many systems over-use male forms. Given a Spanish sentence that does not have a subject pronoun, like in example (8) above, many NMT systems will output a male subject pronoun in English by default. NMT developers are seeking solutions to this problem. Some systems now output both male and female pronouns and let users choose the one they prefer. Other steps that users can take to get the best out of a given NMT system are addressed in Chaptes 4 and 5 of this book.

Other problems have less to do with the translations that NMT systems output and more to do with wider environmental and societal concerns: NMT systems take much longer and much more computing power to train than their predecessors and use up vast quantities of energy in the process. They usually require dedicated, expensive hardware in the form of graphical processing units. They also need massive quantities of training data, which are not available for every language pair.

Improvements in the technology have also led some people to question the wisdom of learning foreign languages: if a machine can translate anything anyone else says or writes in a foreign language into your language, why go to all the trouble of learning their language? Such arguments are based on a very limited understanding of the benefits of second or foreign language learning, however, and ignore the fact that machine translation is viable for only a small number of the world's languages. They also tend to see machine translation as being in competition with language learning, rather than possibly being an aid in the process. Chapters 1, 6 and 9 of this book have more to say on the broader ethical and societal issues raised by the use of machine translation in language learning and other aspects of our lives.

8 Systems, engines and custom NMT

In this chapter so far, we have attempted to explain in a very general way what translation is, what machine translation is and how different types of machine translation work. We draw the chapter to a close with some brief comments on particular machine translation systems and the related concept of a machine translation engine.

In common usage, a machine translation *system* often refers to a machine translation product or service made available by a single supplier or developer. Google Translate is thus understood as Google's machine translation system; while Microsoft has a system called Microsoft Translator. These systems are accessible as services across various platforms. A user might install Google Translate as an app on their mobile phone, for example, or simply use Google Translate on the web, having accessed it using their web browser. They might also access it through an *API* (for "application programming interface") in a third party's software.[20]

---

[20]We use Google Translate here simply because it is probably the most familiar machine translation service. All Big Tech companies offer machine translation "solutions" of one kind or another, as do a whole host of specialist machine translation providers.

In relatively specialized contexts, for example in research papers or professional translation environments, people often talk about machine translation "engines". An *engine* in such contexts is basically a machine translation program (or even a "model") that has been trained to deal with a particular language pair and, often, domain or genre. A commercial machine translation company may, for example, offer its customers access to an English-French engine that was trained on a parallel corpus of financial statements; or a Chinese-German engine that has been trained using only medical texts. Customers might even be able to build or customize their own machine translation engines, using their own data. This kind of service was pioneered by companies like KantanAI.[21] Custom machine translation is discussed in greater depth by Ramírez-Sánchez (2022 [this volume]), and the MultiTraiNMT project has developed a bespoke pedagogical interface that allows students to train their own NMT engines.[22]

9 Four last things you need to know about machine translation

Many readers are likely to use only free, online machine translation and so will encounter only generic engines built for the language pair that interests them. But even these readers should be interested to learn that:

- different systems may output different translations;

- different engines in the same system may output different translations;

- a single system may output different translations for the same input depending on the co-text;

- a single system's outputs may change over time.

For example, at the time of writing, DeepL's French-to-British English engine outputs (16) for the sentence in (15), where the French expression *mon petit doigt me dit* (literally 'my little finger tells me') is used to mean something like 'I have a hunch' or 'someone I won't name has told me'. As the reader will note, the British English translation in (16) uses an entirely appropriate figurative expression with a similar meaning.

(15)   Mon petit doigt me dit que tu es marié.

---

[21]https://www.kantanai.io
[22]http://www.multitrainmt.eu

(16) A little birdie tells me that you are married. (DeepL UK)

Google Translate, on the other hand, outputs the inappropriately literal translation in (17).

(17) My little finger tells me you're married. (Google Translate)

Also at the time of writing, DeepL's French-to-American English engine outputs (18) but if the sentence is changed by a single word as in (19), then DeepL's French-to-American English engine performs much better, as seen in (20).

(18) My little finger tells me that you are married. (DeepL US)

(19) Mon petit doigt me dit que tu es parti.

(20) A little birdie tells me that you've left. (DeepL US)

By the time the reader reads this, however, the outputs of both systems may have changed completely, as models are retrained and users correct faulty outputs.

10 Conclusions

In one way, NMT is just the latest in a line of technologies designed to automate translation, albeit one that has risen to prominence remarkably quickly. Its success could lead to policy makers and ordinary citizens questioning the value of learning foreign languages or training human translators. But such positions would ignore the fact that NMT still relies on human translations or at least translations validated by humans as training data. And because NMT, like other types of machine translation, is not invincible, its outputs still need to be evaluated and sometimes improved by people who can understand both source and target texts. There is also a pressing need for machine translation literacy among even casual users of the technology, so that they do not suffer unnecessarily because of ignorance of how the technology works. Given the right conditions, NMT can be a vital pillar in the promotion and maintenance of multilingualism, alongside language learning and continued translation done or overseen by humans. The rest of this book is dedicated to creating those conditions.

References

Baker, Mona & Gabriela Saldanha (eds.). 2020. *The Routledge encyclopedia of translation studies*. 3rd edition. London/New York: Routledge.

Bao, Guangsheng, Yue Zhang, Zhiyang Teng, Boxing Chen & Weihua Luo. 2021. G-transformer for document-level machine translation. In *Proceedings of the 59th annual meeting of the Association for Computational Linguistics and the 11th international joint conference on Natural Language Processing (volume 1: long papers)*, 3442–3455. Association for Computational Linguistics. https://aclanthology.org/2021.acl-long.267.

Bentivogli, Luisa, Arianna Bisazza, Mauro Cettolo & Marcello Federico. 2016. Neural versus Phrase-Based Machine Translation quality: A case study. In *EMNLP 2016*. arXiv:1608.04631v1.

Bowker, Lynne & Jairo Buitrago Ciro. 2019. *Machine translation and global research*. Bingley: Emerald Publishing.

Carré, Alice, Dorothy Kenny, Caroline Rossi, Pilar Sánchez-Gijón & Olga Torres-Hostench. 2022. Machine translation for language learners. In Dorothy Kenny (ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence*, 187–207. Berlin: Language Science Press. DOI: 10.5281/zenodo.6760024.

Caswell, Isaac. 2022. *Google Translate learns 24 new languages*. https://blog.google/products/translate/24-new-languages/.

Forcada, M. L., M. Ginestí-Rosell, J. Nordfalk, J. O'Regan, S. Ortiz-Rojas, J. A. Pérez-Ortiz, F. Sánchez-Martínez, G. Ramírez-Sánchez & F. M. Tyers. 2011. Apertium: A free/open-source platform for rule-based machine translation. *Machine Translation* 24(1). 1–18.

Forcada, Mikel. 2017. Making sense of neural translation. *Translation Spaces* 6(2). 291–309.

Goodfellow, Ian, Yoshua Bengio & Aaron Courville. 2016. *Deep learning*. Cambridge, MA: MIT Press.

Hutchins, John (ed.). 2000. *Early Years in Machine Translation. Memoirs and Biographies of Pioneers*. Amsterdam/Philadephia: John Benjamins.

Jakobson, Roman. 1959. On linguistic aspects of translation. In Reuben A. Brower (ed.), *On Translation*, 232–239. Cambridge, MA: Harvard University Press.

Johnson, Joseph. 2021. *Worldwide digital population as of January 2021*. statista.com/statistics/617136/digital-population-worldwide/.

Joscelyne, A. 1998. AltaVista translates in real time. *Language International* 10(1). 6–7.

Koehn, Philipp. 2005. Europarl: A parallel corpus for Statistical Machine Translation. In *Proceedings of Machine Translation Summit X*, 79–86. https://aclanthology.org/2005.mtsummit-papers.11.

Koehn, Philipp. 2010. *Statistical Machine Translation*. Cambridge: Cambridge University Press.

Koehn, Philipp. 2020. *Neural Machine Translation.* Cambridge: Cambridge University Press.

Marking, Marion. 2016. *Facebook says Statistical Machine Translation has reached end of life.* https://slator.com/facebook-says-statistical-machine-translation-has-reached-end-of-life.

Moorkens, Joss. 2012. *Measuring consistency in translation memories: A mixed-methods case study.* Unpublished PhD thesis. Dublin: Dublin City University. https://doras.dcu.ie/17332/.

Moorkens, Joss. 2022. Ethics and machine translation. In Dorothy Kenny (ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence*, 121–140. Berlin: Language Science Press. DOI: 10.5281/zenodo.6759984.

Nurminen, Mary & Maarit Koponen. 2020. Machine translation and fair access to information. *Translation Spaces* 9(1). 150–169.

Pérez-Ortiz, Juan Antonio, Mikel L. Forcada & Felipe Sánchez-Martínez. 2022. How neural machine translation works. In Dorothy Kenny (ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence*, 141–164. Berlin: Language Science Press. DOI: 10.5281/zenodo.6760020.

Pym, Anthony. 2010. *Translation and text transfer.* Tarragona: Intercultural Studies Group.

Ramírez-Sánchez, Gema. 2022. Custom machine translation. In Dorothy Kenny (ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence*, 165–186. Berlin: Language Science Press. DOI: 10.5281/zenodo.6760022.

Rossi, Caroline & Alice Carré. 2022. How to choose a suitable neural machine translation solution: Evaluation of MT quality. In Dorothy Kenny (ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence*, 51–79. Berlin: Language Science Press. DOI: 10.5281/zenodo.6759978.

Slobin, Dan. 2003. Language and thought online: Cognitive consequences of linguistic relativity. In Dedre Gentner & Susan Goldin-Meadow (eds.), *Language in mind: Advances in the investigation of language and thought*, 157–191. Cambridge, MA: MIT Press.

Tammet, Daniel. 2006. *Born on a Blue Day.* London: Hodder & Stoughton.

Tammet, Daniel. 2009. *Je suis né un jour bleu.* Paris: J'ai Lu.

Toral, Antonio, Sheila Castilho, Ke Hu & Andy Way. 2018. Attaining the unattainable? Reassessing claims of human parity in neural machine translation. In *Proceedings of the third conference on machine translation: Research papers*, 113–123. Brussels: Association for Computational Linguistics. DOI: 10.18653/v1/W18-6312. https://aclanthology.org/W18-6312.

Turovsky, Barak. 2016. *Ten years of Google Translate*. https://www.blog.google/products/translate/ten-years-of-google-translate/.

Vashishth, Shikhar, Shyam Upadhyay, Gaurav Singh Tomar & Manaal Faruqui. 2019. Attention interpretability across NLP tasks. *CoRR* abs/1909.11218. http://arxiv.org/abs/1909.11218.