OPEN FORUM

Trust, understanding, and machine translation: the task of translation and the responsibility of the translator

Melvin Chen[1,2]

Abstract

Could translation be fully automated? We must first acknowledge the complexity, ambiguity, and diversity of natural languages. These aspects of natural languages, when combined with a particular dilemma known as the computational dilemma, appear to imply that the machine translator faces certain obstacles that a human translator has already managed to overcome. At the same time, science has not yet solved the problem of how human brains process natural languages and how human beings come to acquire natural language understanding. We will then distinguish between the task of translation and the responsibility of the translator. Thereafter, we will conduct a survey of the methods of machine translation (viz. RBMT, SMT, NMT, foundation models or large language models). These methods will then be critically evaluated both in general and relative to Bar-Hillel's hypothesis about the impossibility of fully automatic, high-quality machine translation (FAHQMT). Some concluding remarks will be made about the scope, prospects, and limits of machine translation.

Keywords Machine translation · Task · Automation · Trust · Understanding · Responsibility

1 The task of translation

The task of translation (human or machine) typically involves mapping n linguistic units (e.g. words, phrases, sentences) from a source language S to a target language T, where $n \in \mathbb{N}$ (or $\mathbb{Z}^+$). Furthermore, the task of translation will be successfully completed if and only if two conditions are satisfied:

C1: The meaning of the source string (or n linguistic units from S) is sufficiently well-conveyed in the target string[1];
C2: The target string is grammatically well-formed relative to the syntax of T.

One could easily be lulled into the misconception that the task of translation (human or machine) is relatively straightforward, especially when one considers the facility and natural ease with which children normally acquire a language. Nothing could be further from the truth: natural languages are among the most complex phenomena in the universe (Halliday 2003). The complexity of natural language is a function of its stubborn irregularities that resist law-like generalizations, its richness of expressive potential, its maddening ambiguity, and an endless combination of its linguistic units (Scott 2018). This complexity is compounded by the highly complex mapping that must be made across various levels of linguistic representation (viz. lexical syntactic, semantic, etc) whenever analysis is undertaken (see Fig. 1).

The complexity and ambiguity of natural languages were recognized on behalf of the natural language processing (NLP) branch of AI research quite early on.[2] Furthermore, named entity recognition, word sense disambiguation, and
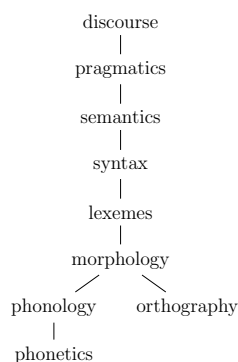

✉ Melvin Chen
melvinchen@ntu.edu.sg

1   School of Humanities, Nanyang Technological University, 48 Nanyang Avenue, Singapore 639818, Singapore

2   Campus for Research Excellence and Technological Enterprise (CREATE), CREATE Way 12-02 CREATE Tower, Singapore 138602, Singapore

[1] Semantic equivalence between the source and target strings, although an ideal, is not required relative to C1. Linguists often argue that, even for human translators, only simple cases show a semantic equivalence between the source and target strings. I am grateful to an anonymous reviewer for having pointed this out to me.

[2] See Council (1966) for the ALPAC (Automatic Language Processing Advisory Committee) report on the prospects of computational linguistics and machine translation.

discourse
|
pragmatics
|
semantics
|
syntax
|
lexemes
|
morphology
 /   \
phonology   orthography
|
phonetics

Fig. 1 Levels of linguistic representation

sentiment analysis are examples of NLP tasks that typically deal with one and only one natural language. Machine translation, on the other hand, has to deal with a language pair (S and T): to the complexity of S, the complexity of T, and the complexity of the interface between S and T must be added the diversity of language pairs.

Consider the scenario in which there are n natural languages in the world in Fig. 2:

It has been estimated by Ethnologue, an annual reference publication that provides information about the living languages of the world, that there are more than 7,000 spoken languages in the world.[3] Where $n = 7,000$, it may be inferred from Fig. 2 that there will be at least 24,496,500 (or $\frac{7000 \times 6999}{2}$) language pairs to be accounted for. This can only accentuate the challenge of successfully completing the task of translation across any one of nearly 25 million language pairs.

2 The computational dilemma

From the complexity, richness, irregularity, ambiguity, and diversity of natural languages, it by no means follows that the task of translation will be an impossible one. On the other hand, these aspects of natural languages, when combined with a particular dilemma known as the computational dilemma, appear to imply that the machine translator faces certain obstacles that a human translator has already managed to overcome.

According to the computational dilemma, any translator (human or machine) assigned with the task of mapping n linguistic units from a source language S to a target language T must embrace one of the following two horns (Barreiro et al. 2011):

H1: Increase the knowledge base to deal with complexity, ambiguity, and diversity of S and T. The more comprehensive the knowledge base, however, the greater the degree of complexity, the more pressing the demand for resources, and the less palatable the environmental consequences;

H2: Reduce the knowledge base to avoid complexity. The smaller the knowledge base, however, the weaker the power of disambiguation.

Humans do not embrace H1 indefinitely: the upper limit in terms of the amount of information that can be accumulated by the average adult human brain has been estimated at around 2.5 petabytes.[4]

Lesions to certain parts of the brain, named after Pierre Paul Broca (1861) and Carl Wernicke (1970) respectively, have resulted in a language disorder known as aphasia. As a result, these areas of the brain — known as Broca's area in the frontal lobe and Wernicke's area in the temporal lobe — have been associated respectively with language production and language comprehension (Fig. 4). Other parts of the brain that have been implicated in natural language processing by humans include the perisylvian cortex, the basal ganglia, and the hippocampus (Duff and Brown-Schmidt 2012). Notwithstanding the progress that has been made, no universal theory has emerged among neuroscientists concerning how the human brain handles natural language (Scott 2018).

Fig. 2 Number of unordered language pairs with n languages

Language (l_i)	l_1	l_2	$\cdots$	l_{n-2}	l_{n-1}	l_n
Language pairs	l_1-l_n	l_2-l_n		l_{n-2}-l_n	l_{n-1}-l_n	$\langle$null$\rangle$
	l_1-l_{n-1}	l_2-l_{n-1}		l_{n-2}-l_{n-1}		
	$\vdots$	$\vdots$	$\cdots$			
	l_1-l_3	l_2-l_3				
	l_1-l_2					
Number of pairs	$(n-1)$	$(n-2)$	$\cdots$	2	1	0
Total number of pairs (N)	$N = 1 + 2 + \cdots + (n-2) + (n-1) = \frac{n \times (n-1)}{2}$					

Unit of data	Description	Approximation
1 bit (binary digit)	1 (true) or 0 (false)	
1 byte	8 bits	
1 kilobyte (KB)	1,024 bytes	1×10^3 bytes
1 megabyte (MB)	$1 \times 1,024^2$ bytes	1×10^6 bytes
1 gigabyte (GB)	$1 \times 1,024^3$ bytes	1×10^9 bytes
1 terabyte (TB)	$1 \times 1,024^4$ bytes	1×10^{12} bytes
1 petabyte (PB)	$1 \times 1,024^5$ bytes	1×10^{15} bytes
1 exabyte (EB)	$1 \times 1,024^6$ bytes	1×10^{18} bytes

**Fig. 3** Units of memory or data storage capacity

While human translators offer an excellent example of how this computational dilemma, though forbidding, is not insuperable, it is not clear how this achievement is effected.

3 The responsibility of the translator

The task of translation, as described in §1, is however no ordinary task. It is distinct from the task of processing financial transactions and the task of processing the components of a product and constructing that product stepwise along a production line. We do not have any issues with ATMs handling deposits, cash withdrawals, and transfers between accounts, nor do we generally have any issues with automated manufacturing techniques being employed in industries. The task of translation involves handling natural languages instead of financial transactions or components of a product.[5] According to Charles Francis Hockett (1967; 1959), a set of features distinguishes natural languages from other forms of animal communication.[6] These features do not merely indicate that natural languages are complex: they imply in addition that the use of natural languages is both unique to the human species and a central aspect of human experience. Of these features that allow us to distinguish between natural languages and other forms of animal communication, Hockett considers duality of pattern to be the most distinctive: all meaningful elements of a natural language (viz. morphemes, words, sentences, etc) are composed from a limited inventory of meaningless elements (viz. phonemes) (Gair 2006).

[5] By definition, a natural language refers to any language that has evolved naturally in human beings through use and repetition.

[6] These design features include the following: displacement (the ability to refer to the past, the future, and what might not correspond to reality), productivity (openness, creativity, and the reflexivity of language), cultural transmission (the ability to learn languages from elders in the community), duality of patterning (the use of combinations of a small number of discrete speech sounds to produce a large number of meaningful words and sentences).
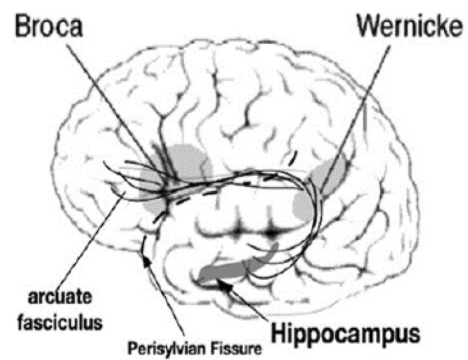
Fig. 4 Language areas of the brain (Scott 2018), adapted from National Institute of Health Publication 97-4257

Given the unique and species-specific nature of natural languages, the task of translation (human or machine) must be approached with care. Unlike the tasks of processing financial transactions and constructing products along a production line, the task of translation is a sensitive one. This implies that certain normative expectations arise with machine translation programs that do not typically arise with respect to ATMs and automated manufacturing systems. It is plausible to assert that we cannot trust the linguistic behaviour of translators (human or machine) unless they understand the source and target languages that are involved in the task of translation. It is therefore the responsibility of translators to ensure both that they understand the source and target languages S and T and that the task of translation is successfully completed.

4 The metaphysics & epistemology of understanding

I have distinguished between the task of translation (§1) and the responsibility of the translator (§3). Furthermore and given the unique and sensitive nature of the former, the latter will have as a necessary prerequisite an understanding of the source and target languages S and T involved in the task of translation. Whether translation can be automated depends chiefly on whether the two conditions C1–C2 of the task of translation (§1) can be satisfied by automated procedures. On the other hand, whether translation should be automated depends chiefly on whether the normative expectations that we have of human translators (viz. that they understand the source and target languages involved in the task of translation) can be satisfied by their fully automated alternatives. Whether and how machine translation programs could come to understand natural languages in the way that human beings do will determine whether translation should be automated, however successful these programs might be in the task of translation. The metaphysical

account of understanding that we endorse will influence what we take understanding ultimately to be, whether the translator (human or machine) demonstrates a requisite level of natural language understanding, and whether translation ought to be automated.

Although the metaphysics of understanding and the epistemology of understanding constitute two distinct branches of the philosophy of understanding, there is a natural tendency to conflate them (Bommasani et al. 2021). Firmer foundations in the philosophy of understanding will allow us to sharpen what we mean by 'understanding' when we claim that a key part of the responsibility of the translator involves understanding the source and target languages S and T. What is it in virtue of which we can say of translators that they possess natural language understanding (i.e., the correct grasp of the meanings of the relevant linguistic phenomena)? This is the central question in the metaphysics of understanding. A response to this question will typically be supported by a theory of meaning (e.g. internalist, referentialist, pragmatist). According to an internalist theory of meaning, the meanings of linguistic phenomena are equivalent to internal mental structures, arrived at in a compositional fashion from the meanings of words and their syntactic arrangement. This implies that a translator who understands the source and target languages S and T will map the n linguistic units from S to a meaning (e.g. a concept, a representation, or some other internal object of understanding), before generating a target string that has the same or a similar meaning.

According to a referential theory of meaning, on the other hand, words have external referents and declarative sentences are truth-apt (i.e., they could be true or false). Translators who understand the source and target languages S and T are capable of determining whether sentences in languages S and T are true relative to a context. A translator who understands the source and target languages S and T might map from a word or phrase in S to an external referent, before mapping from this external referent to its matching word or phrase in T.[7] According to a pragmatist theory of meaning, understanding requires neither the appropriate internal representations (e.g. mental structures as identified by internalists) nor the appropriate relationship with truth and reference (e.g. external referents as identified by referentialists) (Wittgenstein 1953). Rather, the pragmatist is concerned with whether agents are disposed to use natural languages in the appropriate manner. What matters here is the exhibition of the appropriate behavioral dispositions. One of the most famous pragmatist proposals in the context of AI and NLP research involves replacing the question of

whether machines can think with the question of whether a machine can be constructed to pass a test for human intelligence satisfactorily (Turing 1950).

Whereas the metaphysics of understanding is concerned with what it would mean for an agent to achieve the relevant sort of understanding, the epistemology of understanding is concerned with how we could come to determine that an agent has achieved the relevant sort of understanding. If the pragmatist theory of meaning is correct, then success conditions for understanding may be identified with the manifestation of appropriate linguistic behaviours and behavioral dispositions. In the context of machine translation, a behavioral test for the natural language understanding possessed by human translators may involve a set of translation tasks. A machine translation program that is capable of satisfying conditions C1–C2 for each of these translation tasks will have manifested the appropriate linguistic behaviours and this will be a sufficient indicator that it possesses the requisite understanding. At the same time, whenever chatbots or conversational programs pass some version of the Turing test, their behavioral dispositions typically do not function as an apodictic indicator that they possess the requisite conversational intelligence. Instead of identifying the actual behaviour of these chatbots or conversational programs with the target or desired behaviour, AI researchers tend to regard the test-passing dispositions of these chatbots (however reliable they might be) as a symptom that the test itself is flawed. All things considered, the pragmatist theory of meaning does not appear to offer the correct metaphysical support with respect to what understanding ultimately is or what it might entail.

The reservations that arise when chatbots pass the Turing test may be extended to machine translation programs when they successfully complete various translation tasks. These tests and tasks are imperfect and defeasible means of assessing whether understanding has been achieved. On the metaphysical front, our gold standard probably remains either internalism, referentialism, some combination of both, or some alternative theory of meaning that is neither internalist, referentialist, nor pragmatist. The epistemological implication here is that we need, if we are to determine that machine translation programs have achieved the relevant sort of understanding, to find some reliable means of tracking their inner workings and dynamics and determining whether they are appropriate. To the extent that these means remain absent or incomplete, we cannot say with confidence that these machine translation programs, however successful they might be at the task of translation, are able to secure our trust in virtue of their possessing the requisite level of natural language understanding. This implies that we have good epistemological grounds to deny that translation should be fully automated.

---

[7] A useful introduction to internalist and referential theories of meaning may be found in Elbourne (2011).
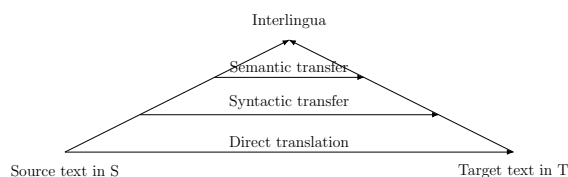
Fig. 5 Vauquois triangle for RBMT

5 Machine translation methods

Machine translation researchers have not waited for neurolinguistics to solve the problem of how human translators cope with the computational dilemma §2 (nor should they). In addition, machine translation researchers have not waited for philosophers to deliver the correct metaphysical account of understanding (e.g. internalist, referentialist) and the epistemological means for determining whether the gold standard of understanding has been achieved. Instead, they have applied themselves conscientiously to the development of various methods of machine translation: rule-based machine translation (hereafter: RBMT), statistical machine translation (hereafter: SMT), and (more recently) neural machine translation (hereafter: NMT).

5.1 RBMT

The Vauquois triangle is a hierarchical model that allows us to visualize various machine translations approaches (Vauquois 1968). Figure 5 depicts a Vauquois triangle relative to RBMT:

Direct translation (or dictionary-based translation) is an example of RBMT that relies on a direct word-for-word translation from S into T for all words in the *n* sentences to be mapped in the translation task. Here, the complexity that arises from mapping across different levels of linguistic representation (Fig. 1) is avoided as analysis and syntactic reorganization are kept to a minimum.[8]

Transfer-based machine translation is another example of RBMT: it relies on an analysis of the grammatical structure of each of the *n* sentences in S, the application of rules to transfer to a suitable structure for translation in T based on knowledge of the differences between S and T, and a generation of the target output in T relative to this suitable

[8] Direct translation may rely on morphological analysis (analyzing the structure and parts of words) or lemmatization (converting words to their lemmas or basic dictionary forms).

[9] Tree-based parsing is a process in which a flat input sentence like 'This boy walks quickly' can be converted into a hierarchical structure. An online resource for sentence-parsing is available: https://www.link.cs.cmu.edu/link/submit-sentence-4.html.
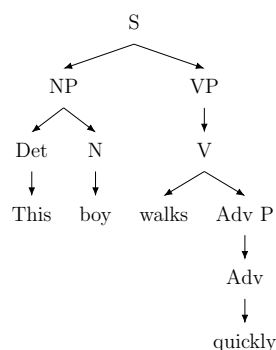

Fig. 6 Parse tree diagram for the sentence 'This boy walks quickly'

structure.[9] Figure 6 offers an example of a parse tree structure for a sentence in English:

Where S and T are highly similar (e.g. Czech and Slovak), only syntactic transfer may be needed. Conversely, where S and T are quite different (e.g. Vietnamese and English), semantic transfer may be needed on top of syntactic transfer. The interlingual approach is yet another example of RBMT: it analyses the *n* sentences in S, represents them as an interlingua, and then generates the target output in T on the basis of this interlingual representation. The interlingua is an intermediate and abstract representation of meaning that facilitates translation between S and T. Semantico-syntactic Abstraction Language (hereafter: SAL), a representation employed by the Logos model, is a famous example of an interlingua: SAL is a second-order taxonomic language (similar to hypernyms) to which words in a natural language map.[10] As a language, SAL boasts more than 1,000 words or elements, organized in a hierarchical taxonomy comprising supersets, sets, and subsets, distributed over all parts of speech (Barreiro et al. 2011).

Direct translation restricts its analysis to the levels of morphology and lexemes: this allows it to avoid the complexity associated with mapping across various levels of linguistic representation. This embrace of H2 (or the second horn) of the computational dilemma (§2), however, has its costs: direct translation is less able to resolve ambiguity in meaning and cope with the complexity and diversity of S and T. This entails that the quality of translation output in T will be limited. The other examples of RBMT embrace H1 instead of H2: they typically add to the store of rules in the knowledge base. While a greater number of rules across various levels of linguistic representation will enhance the power of disambiguation and increase the quality of translation, it also implies a larger and more unwieldy knowledge base. As developers of rule-based systems work exceptions into generalized rules, they will have to address exceptions to

[10] Just as 'cat' has for its hypernym 'animal' or 'pet', a word like 'table' would have for its SAL representation the more abstract 'supporting surface'.

Fig. 7 English translations of 'Haus' in the German-English bilingual corpus C (Koehn 2009: p. 84)

English translation t of 'Haus' (s_2)	Count	Estimation of translation probability or $P(s_2|t_j)$
house	8,000	$P(s_2|t_j) = \frac{8000}{10000} = 0.8$ if $t_j = $ house
building	1,600	$P(s_2|t_j) = \frac{1600}{10000} = 0.16$ if $t_j = $ building
home	200	$P(s_2|t_j) = \frac{200}{10000} = 0.02$ if $t_j = $ home
household	150	$P(s_2|t_j) = \frac{150}{10000} = 0.015$ if $t_j = $ household
shell	50	$P(s_2|t_j) = \frac{50}{10000} = 0.8$ if $t_j = $ shell

exceptions, exceptions to these exceptions to exceptions, etc. Furthermore, as the sequence of procedural logic governing these rule-based systems becomes increasingly complex, it may become difficult or even impossible for the developers to keep track of and manage the logic, the likelihood of logical inconsistencies or contradictions arising will increase, and some logic that might be written to resolve an issue might undo the earlier resolution of another issue (Scott 2018).

5.2 SMT

Given the complexity that arises from an increasing number of rules in rule-based systems (§5.1), might it be possible to avoid rules altogether and rely instead on the statistical mining of raw language? SMT, an approach to machine translation based on a Bayesian approach to inference, believes that this can be done (Brown et al. 1988, 1990). Here is Bayes' theorem (BT) (Bayes 1763)[11]

$$(BT)\ P(X|Y) = \frac{P(Y|X)\ \times\ P(X)}{P(Y)}$$

Each text is translated in accordance with a probability distribution $P(T|S)$ that a string T in target language T is the translation of a string S of n sentences in source language S. In accordance with Bayes' theorem:

$$P(T|S) = \frac{P(S|T)\ \times\ P(T)}{P(S)}$$

$P(T|S)$ is the probability that a translator will produce T when presented with S. Furthermore, $P(S)$ remains constant for each string T that is under consideration and may therefore be regarded as independent of T. The following may be inferred:

$$P(T|S) \propto P(S|T)\ \times\ P(T)$$

---

[11] The proof of BT runs as follows:

1. $P(X|Y) = \frac{P(X \cap Y)}{P(Y)}$

2. $P(Y|X) = \frac{P(X \cap Y)}{P(X)}$

3. $\therefore P(X \cap Y) = P(Y|X) \times P(X)$ — from 2.

4. $\therefore LHS = P(X|Y) = \frac{P(Y|X) \times P(X)}{P(Y)} = RHS$ — from 1 & 3 (QED).

The translation model $P(S|T)$ assigns the probability that S is a translation of T and ensures accuracy of translation. The language model $P(T)$ assigns the probability of seeing the string T in the target language and ensures fluency of translation. Under this Bayesian-inspired SMT approach, the desideratum is the most likely translation T for the some S. The equation to find the most probable T may be represented as follows:

$$\hat{T} = \underset{T}{\mathrm{argmax}}\ (P(T) \times P(S|T)) = \underset{T}{\mathrm{argmax}}\ P(T|S)$$

The SMT approach entails that the best translation of a string S in S is the translation T in T that is the most probable. Ideally, T is the most accurate translation (i.e., the value for $P(S|T)$ assigned by the translation model is the highest) and the most fluent translation in the target language (i.e., the value $P(T)$ assigned by the language model is the highest).

Consider the German sentence 'Das Haus ist klein'. This string S comprises four words s_1, s_2, s_3, s_4 appearing in a specific order. Suppose further that the following English translations of 'Haus' (s_2) may be found in the German-English bilingual corpus C with the following frequencies (Fig. 7):

Proceeding in a word-by-word fashion, the most probable translation for 'Haus' (s_2) would be 'house' (t_j, where $P(s_2|t_j) = 0.8$). Next, considerations about how each target word t_j is aligned with its associated source word s_i will ensue. Alignment may be formalized with an alignment function: each source word will be mapped at position i to a target word at position j with the alignment function $a : i \mapsto j$.

Relative to the alignment of S (in German) and T (in English) in Fig. 8, the alignment function that will yield the mapping $a : \{1 \mapsto 1, 2 \mapsto 2, 3 \mapsto 3, 4 \mapsto 4\}$. This is a relatively straightforward alignment and it simply means that the German source words and their English target counterparts are in exactly the same order. All other things being equal and relative to an SMT approach, the anticipated translation T will be 'The house is small'.

By contrast, suppose that our source string S (in English) is 'You will find him in the garden' and its target string T (in French) is 'Tu peux le trouver dans le jardin'. Figure 9 represents the associated diagram for alignment:

Relative to the alignment of input English text and output French text in Fig. 9, the

0	1	2	3	4
⟨null⟩	Das	Haus	ist	klein
⟨null⟩				
0	1	2	3	4
⟨null⟩	The	house	is	small
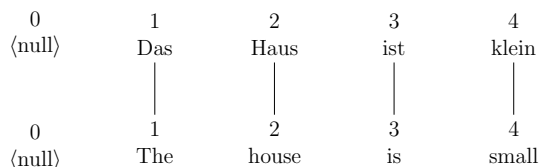
Fig. 8 Alignment for 'Das Haus ist klein' (*S*) and its target text *T* in English, adapted from Koehn (2009: p. 84)

alignment function may be represented as $a : \{1 \mapsto 1, 2 \mapsto 2, 3 \mapsto 4, 4 \mapsto 3, 5 \mapsto 5, 6 \mapsto 6, 7 \mapsto 7\}$. While SMT avoids the complexity that is associated with an ever-growing number of rules in rule-based systems, its embrace of H1 equally entails a greater degree of complexity. The knowledge base in this instance would be the bilingual corpus C relative to which the values for statistical frequencies and conditional probabilities are computed. The larger the size of C, the more likely it will be that both S and T are captured in all their respective complexity, richness, irregularity, and diversity and the more representative will the statistical frequencies and conditional probabilities be. Simultaneously, a larger C will entail a greater demand for resources and a higher toll on the environment.

Corpus-based approaches like SMT are designed to prioritize accuracy in translation over lexical range and diversity of output. These approaches may result in a greater levelling out of texts, standardization and normalization, loss of lexical richness, and language impoverishment than human translation (Klebanov and Flor 2013; Vanmassenhove et al. 2019b). Furthermore, the reliance on statistical dependencies may lead to certain biases in the dataset being replicated (Vanmassenhove et al. 2019a). In other words, the reliance on C is a double-edged sword: while a sufficiently representative C is a good guide to high-quality translation, an uncritical reliance on C may equally lower the quality of translation.
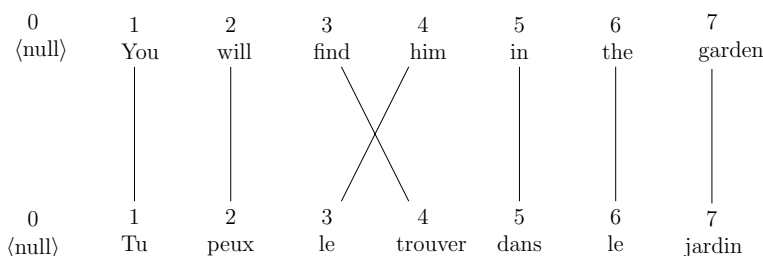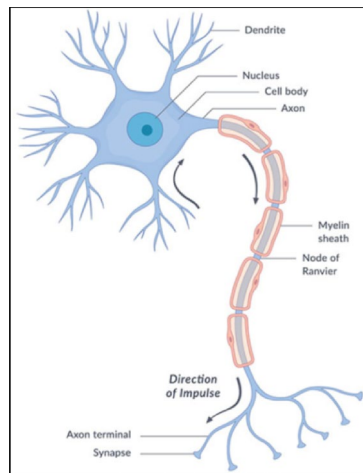
5.3 NMT

NMT relies on artificial neural networks (ANNs) to learn the mapping from *S* (text in the source language S) to *T* (output in the target language T) (Bahdanau et al. 2014; Sutskever et al. 2014). State-of-the-art machine translation systems rely on NMT and they include Baidu (whose transition from SMT to NMT began in 2015), Google (whose transition from SMT to NMT began in 2016), Microsoft (whose transition from SMT to NMT began in 2016), and Systran (whose transition to NMT also began in 2016).
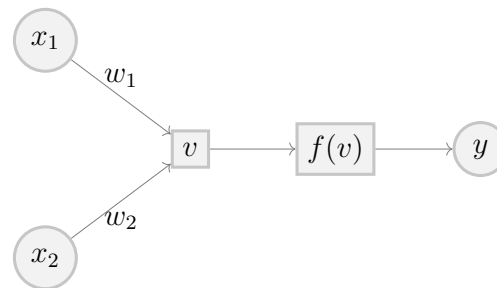
Each artificial neural network (ANN) is composed of a collection of connected nodes known as artificial neurons. Artificial neurons are loosely modelled after biological neurons in the brain, as illustrated by Fig. 10:

Relative to Fig. 10a, each neuronal cell contains three main parts: dendrites, a cell body, and an axon.[12] The neuronal cell is an electrically excitable cell that takes up, processes, and transmits information through electrical and chemical signals. Electrical signals may first be received by the dendrites and brought to the cell body. These signals could be conducted away from the cell body by the axon. The electrical signals are eventually transmitted by the axon terminals of the presynaptic cell and received by the dendrite receptors of the postsynaptic cell. Relative to Fig. 10b, x_1 and x_2 denote inputs and w_1 and w_2 denote their associated weights. v is equivalent to $(x_1 \times w_1) + (x_2 \times w_2)$. $f(v)$ and T respectively denote the thresholding function and the threshold, such that $f(v) = 1$ when $v \geq T$ or $f(v) = 0$ when $v < T$. Last but not least, y denotes the output such that $y = f(v)$ (i.e., 0 or 1). Artificial neural networks (ANNs) may either have a single layer or multiple layers (preferred), as illustrated in Fig. 11:

Relative to the single-layered perceptron (Fig. 11a), $x_1, x_2, \cdots, x_n$ denote the *n* inputs, $w_1, w_2, \cdots, w_n$ denote their corresponding weights, σ denotes the thresholding function, θ denotes the threshold (or negative bias), and y denotes the output such that

Fig. 9 Alignment for 'You will find him in the garden' (*S*) and its target text *T* in French

0	1	2	3	4	5	6	7
⟨null⟩	You	will	find	him	in	the	garden
⟨null⟩	Tu	peux	le	trouver	dans	le	jardin
0	1	2	3	4	5	6	7

(a) Anatomy of a biological neuron

(b) McCulloch-Pitts model of an artificial neuron with 2 inputs

Fig. 10 Biological versus artificial neurons

$$y = \text{sign}\left(\sum_{i=1}^{n}(x_i \times w_i) - \theta\right)$$

$$= \begin{cases} 1 & \text{if } (\sum_{i=1}^{n}(x_i \times w_i) - \theta) \geq 0 \\ 0 & \text{if } (\sum_{i=1}^{n}(x_i \times w_i) - \theta) < 0 \end{cases}$$

Multi-layered networks (Fig. 11b) are more complex than elementary perceptrons (Fig. 11a), allowing for decision boundaries of greater complexity. Multi-layered ANNs may be feedforward or recurrent. Whereas traditional feedforward ANNs assume that the inputs $x_1, x_2, \cdots, x_n$ and output y are independent of each other, recurrent neural networks or RNNs assume conversely that information from prior inputs $x_1, x_2, \cdots, x_{j-1}$ will influence the current input x_j and output y. Figure 12 illustrates the structure of RNNs:

An RNN (Fig. 12) is an ANN that may be employed with sequential data or time series data. RNNs can be used to addressed temporal problems (e.g., natural language processing, machine translation, speech recognition, etc). It should be noted that NMT typically relies on RNNs. An example from IBM should help to underscore the relevance of RNNs to machine translation.[13] The idiom 'feeling under the weather' is used to express how someone is feeling ill. The words in this idiom must be expressed in exactly that sequence if the idiom is to make any sense. It should therefore not surprise anyone that RNNs need to account for the position of each word in the idiom and use that information to predict the next word in the sequence.

In addition, the three key components in the NMT architecture are the encoder network (RNN), the attention module, and the decoder network (RNN). The RNN-based encoder transforms an input source text S into a list of vectors. Thereafter, the attention module allows the decoder to focus on different regions of S during the decoding process. Finally, the RNN-based decoder produces the translation T one symbol at a time.[14]

The larger and more representative the bilingual corpus C on which NMT is trained, the larger the ANN (typically measured in terms of the number of training parameters or weights relative to which the input data S might be transformed), the better the ability of NMT to generate distributed representations of words as vectors. Like most forms of RBMT (§5.1) and SMT (§5.2), NMT embraces H1 of the computational dilemma and embodies a 'more is better' ethos, both in terms of the size of the training dataset and the size of the ANN. However, the same objections that beset corpus-based approaches like SMT will equally apply to NMT.

In addition, the higher the number of training parameters or weights, the greater the complexity of the entire ANN, the more pressing the demand for resources, and the less palatable the environmental consequences. After they have been trained on a dataset, vectors (or large matrices of real-numbered values) are what decide in favour of a specific choice of words as opposed to some alternative. The

---

[13] https://www.ibm.com/cloud/learn/recurrent-neural-networks

[14] The special end-of-sentence symbol (EOS) is produced at the end of each sentence. For an example of an NMT system running on an attention-based encoder-decoder model, see Google on GitHub: https://google.github.io/seq2seq/ An alternative to the attention-based encoder-decoder model involves the use of convolutional neural networks or CNNs (Kalchbrenner and Blunsom 2013). However, virtually all NMT research has focused on the attention-based encoder-decoder model.
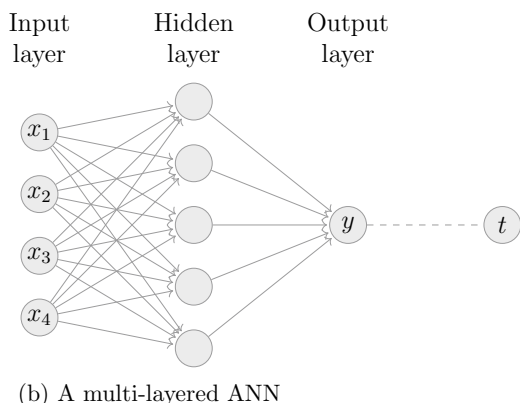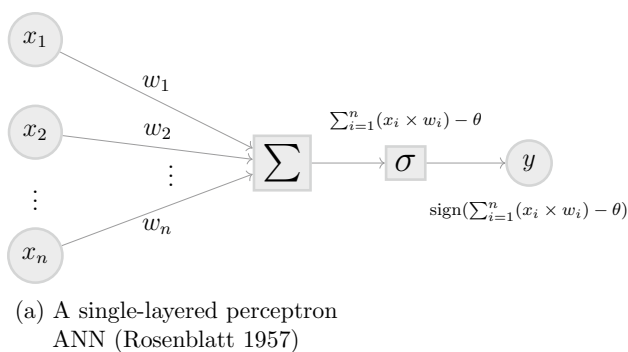
(a) A single-layered perceptron ANN (Rosenblatt 1957)

$$\sum_{i=1}^{n}(x_i \times w_i) - \theta$$

$$\text{sign}\left(\sum_{i=1}^{n}(x_i \times w_i) - \theta\right)$$

(b) A multi-layered ANN

Fig. 11 Single- versus multi-layered ANNs

black-box nature of NMT systems in particular and machine learning in general invites the objection that NMT is inscrutable: we cannot explain how T might ultimately have been derived from S. Furthermore, NMT systems tend to be powerless when confronted with words in S that they have not been trained to deal with. It is easy to construct sentences that have neither been written nor uttered in the history of the world. NMT systems that have been trained on billions of sentences in the bilingual corpora can nonetheless be stumped by these sentences, since they lack the power of generality (Scott 2018).[15]

5.4 Foundation models

Given the amount of hype surrounding its emergence, ChatGPT must be covered in any survey of machine translation methods. ChatGPT is part of an emerging paradigm for building AI systems, based on a general class of models known as foundation models or large language models (Bommasani et al. 2021). A foundation model may be defined as a model that is typically trained on vast amounts of data (generally using self-supervision at scale) and can be adapted or fine-tuned to a wide range of downstream tasks.

---

[15] The same can be said of SMT systems. Rules, which grant the power of generality, are characteristic only of RBMT systems.
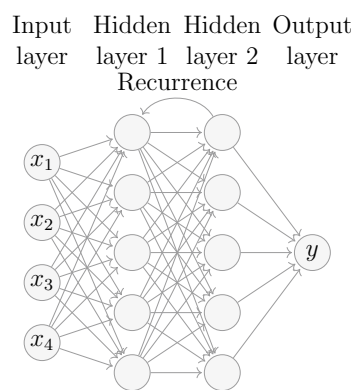

Fig. 12 An RNN

Examples of foundation models include BERT, CLIP, and GPT-3 (Brown et al. 2020; Devlin et al. 2019; Radford et al. 2021). ChatGPT is effectively a proof of concept of foundation models: it is a chatbot developed by OpenAI and powered by OpenAI's GPT large language model.

On a technical level, foundation models are enabled by transfer learning and scale. By transfer learning is meant the ability to take knowledge derived when solving a problem and apply it to solve a different though related problem (Thrun 1998). In the context of deep learning-based AI systems, pretraining is the preferred approach to transfer learning: a model is trained on some surrogate task, before being adapted or fine-tuned to the downstream task of interest. By scale is meant the availability of ever-vaster amounts of training data, improvements in computer hardware, and the emergence of transformer model architecture to improve the context-based processing of unlabelled text across a broad range of tasks and train increasingly more expressive models (Bommasani et al. 2021). Google's transformer model architecture, first developed by Vaswani et al. (2017), underpins most of the foundation models to date. Nonetheless and as things stand, we are still dealing with stochastic parrots or lumbering statistical machines for pattern-matching, rather than entities that are capable of natural language understanding after the manner of human beings (Bender et al. 2021; Chomsky et al. 2023).

6 FAHQMT

During 1958–1959, Yehoshua Bar-Hillel was charged by the U.S. Office of Naval Research to make a critical assessment of machine translation activity in the United States and Great Britain. Where FAHQMT denotes fully automatic, high-quality machine translation, Bar-Hillel ended up defending the hypothesis according to which FAHQMT is impossible in principle (Bar-Hillel 1960). This may be usefully contrasted with attempts to develop models to represent natural language,

store linguistic knowledge, apply the knowledge base to an input, and address complexity effects as the knowledge base grows year after year (Scott 2003).[16] According to Bar-Hillel, the goal of FAHQMT or fully automatic translation of a quality equal to that of a competent human translator is unattainable, not only in the near future but also in principle. Bar-Hillel's proof or demonstration of the truth of his FAHQMT hypothesis makes reference to the following trio of sentences:

> Little John was looking for his toy box. Finally he found it. The box was in the pen.

The sentence-final 'pen' has at least two meanings: a play-pen or enclosure where small children play (MEANING 1) and a writing instrument (MEANING 2). Given the relative sizes of play-pens (MEANING 1 of 'pen'), toy boxes, and writing instruments (MEANING 2 of 'pen'), only MEANING 1 is plausible in this context. The full resolution of ambiguity and discovery of the true meaning of 'pen' relies on general knowledge about the world and a human-like understanding of states of affairs. However, it is impossible to build such general knowledge or human-like understanding into computers. Therefore, neither fully automated nor human-quality translation ought to constitute realistic goals for machine translation researchers.

The contemporary practice of machine translation supports Bar-Hillel's FAHQMT hypothesis. While the rules of RBMT grant a certain power of generality, they are strictly rules across various levels of linguistic representation that are designed to resolve ambiguity. They do not grant machine translation systems the power to acquire general knowledge about the world and a human-like understanding of states of affairs. While SMT and NMT (usage-based) have the ability to learn from usage (i.e., they can learn from having been exposed to bilingual corpora), this is still insufficient to count as general knowledge about the world and a human-like understanding of states of affairs. The lack of this power of generality in statistics- and usage-based machine translation systems is most evident when these systems have to handle out-of-distribution source strings. Hybrid systems might appear to be more promising: the Logos model relies on both a rule-based approach (recall our discussion about its use of SAL representation in §5.1) and a usage-based approach (semantic associations and through the matching of patterns). It may be argued that SAL provides the requisite level of disambiguation relative to Bar-Hillel's example. SAL has MEANING 1 and MEANING 2 stored and then can be used in a rule that constrains the meaning of 'pen' when there is a container ('box'). The container cannot be inside the 'pen' (an instrument to write with). SAL allows for this disambiguation.

Note however that SAL may fare less well when the container is a pen refill, which contains both actual ink and the gel that prevents this ink from flowing out from the refill through the top. Pen refills are containers that are normally located inside pens. Human beings understand that while most containers cannot be inside a pen, the pen refill is an example of a container that is contained within a pen. All things considered, even hybrid systems like the Logos model lack general knowledge about the world and a human-like understanding of states of affairs and flounder with modified versions of Bar-Hillel-style examples.

The appropriate use of natural languages presupposes both a general knowledge and an understanding of the world that RBMT, SMT, NMT, hybrid approaches, and foundation models have been unable to deliver. An example from Bender and Koller (2020) will suffice to illustrate how it remains doubtful whether machine translation programs, however successful they might appear to be in the task of translation, will completely secure the trust of human users. They imagine an agent O (a hyperintelligent deep-sea octopus) that intercepts communications between two human beings A and B speaking a natural language L. This agent O, in virtue of its inhabiting a world starkly different from that of the human beings A and B, does not have the appropriate sort of experiences to ground human utterances.[17] Nonetheless, O might be able to learn from the patterns in the utterances of A and B to the extent that it can pass off as a human being. However, we can still easily envisage situations in which O's inability to ground the natural language in the world of human beings (as required by referentialism) or the internal representations or mental structures of human beings (as required by internalism) will show O up in at least certain circumstances. These circumstances will be circumstances in which it shall be revealed that O does not understand the natural language used by A and B. Given the complexity of the external world and our internal representations of this world, it is doubtful whether any amount of textual data can fully encompass this complexity and the gaps will eventually reveal themselves.

Whatever may be said about agent O will equally apply to machine translation programs, whether they have been

---

[16] Bernard Scott is one of the pioneers of commercial machine translation whose attempt to develop models of this nature in the quest for FAHQMT must be fully acknowledged. I am grateful to an anonymous reviewer for directing me to Scott and his intellectual legacy.

[17] O will therefore not possess understanding in the sense required by referentialism. The point may be extended to incorporate an internalist account of understanding: while machine translation programs may be good at the levels of words or even sentences, they tend to treat discourse as a mere conjunction of sentences. Insofar that machine translation programs will not have the appropriate internal representations of discourse, they will also not possess understanding in the sense required by internalism. It may be argued that if O has SAL knowledge, then it will be able to learn more about the words. However, SAL knowledge is still linguistic knowledge, and SAL-based systems have no idea of the underlying reality that natural language describes. Human users of natural languages regularly rely on common sense (nonlinguistic knowledge) and it remains doubtful whether we can design AI systems with the level of common sense required for natural language understanding.

developed in accordance with RBMT, SMT, NMT, hybrid, or foundation model approaches. As language modelling tasks use only linguistic form as training data, machine translation programs that have been trained on form alone have no *a priori* way to learn meaning and acquire the requisite understanding of natural language. Insofar that they lack the requisite natural language understanding, we can predict that machine translation programs will fail to secure the trust of human users. With respect to ChatGPT, these gaps reveal themselves when it vacillates between coherent and nonsensical responses depending on the exact nature of the prompt being used, generates information that, though grammatically well-formed and correct-sounding, is incorrect, and produces harmful or biased content.

To recapitulate, here are the two conditions that must be satisfied relative to each task of translation (§1):

C1: The meaning of the source string is sufficiently well-conveyed in the target string;
C2: The target string is grammatically well-formed relative to the syntax of T.

However successful machine translation programs might be in the performance of discrete translation tasks, it by no means follows that they have fulfilled the responsibility of a translator. Gaps between the linguistic form of the training data and linguistic meaning (whether relative to the external referents invoked by referentialists or the internal mental structures invoked by internalists) ensure that machine translation programs will lack the requisite level of natural language understanding and fail to secure the trust of human users. In any case, human intervention remains an indispensable component in machine translation, in the form of pre-editing and post-editing.[18] In the pre-editing phase, the human pre-editor will address ambiguities and rearrange the source text in accordance with a standard order in the target language, following instructions available to her in her own language. No prior knowledge of the target language is required from the human pre-editor. In the post-editing phase, the human post-editor will remove errors from the machine translation output, ensuring that the meaning is correct. The post-editor will also attend where possible to stylistic issues. Here, knowledge of the target language is required from the post-editor.

The practices of pre- and post-editing and the central role that they occupy in machine translation suggest that fully automated and intervention-free translation will remain a pipe dream.[19] Human intervention from pre- and post-editors is still required to make the source text amenable to machine translation, correct mistakes in the source text, resolve ambiguity, simplify structures, eradicate mistakes in the machine translation output, attend to stylistic issues, and attain a high level of quality in translation.[20] Human intervention is what ultimately secures trust in processes that rely on machine translation, since responsibility may be attributed to human beings who possess the requisite level of natural language understanding.

At the same time, even if Bar-Hillel's FAHQMT hypothesis is correct and FAHQMT is impossible in principle, it by no means follows that all forms of machine translation are impossible or useless. Machine translation could still be used to produce output of a sufficiently high quality to reduce the level of human post-editing effort required. We may have to start thinking of automation in terms of degrees of automation, as opposed to an all-or-nothing affair. There may be different levels of automation in machine translation. We may also have to concede that machine translation is likely to fare better at certain tasks as opposed to others: the task of translating technical language, for instance, is much more amenable to machine translation than the task of translating a piece of literature or poetry, provided that the technical terms are in the system terminologies (Barreiro et al. 2011). Certain bottlenecks are bound to appear: knowledge about the real world, commonsensical knowledge, cultural knowledge, and even knowledge about certain higher-level aspects of language (e.g. pragmatics, discourse analysis) may be difficult or even impossible to build into AI systems. These bottlenecks will mark the advantage that human translators continue to enjoy over their machine counterparts and the contours of future job descriptions for human pre- and post-editors and future research agendas for machine translation researchers. At the same time, if we wish to have some means of determining whether or to what extent machine translation programs have achieved the relevant sort of understanding, then developing methods for tracking and studying the inner workings and dynamics of machine translation programs may be useful (Sundararajan et al. 2017; Tenney et al. 2019).[21]

---

[18] The concepts of pre-editing and post-editing in MT may be traced to Erwin Reifler (1952), a professor of Chinese Languages and Literature.

[19] At the same time, there is no inconsistency in maintaining that machine translation can still be improved so that less effort is required in the editing task.

[20] Equally, human intervention may be required in the curation of corpora on which AI systems are trained. The corpora may contain a copious amount of noise, unintelligent or meaningless text, and text from ethically questionable sources (e.g. propaganda generators). I am grateful to an anonymous reviewer for having pointed this out to me.

[21] This entails that we will need to reconsider how we make the accuracy-interpretability trade-off in machine translation. Black-box models are more accurate but their inner workings and dynamics are more difficult to understand. By contrast, white-box models are less accurate but their inner workings and dynamics are easier to explain and interpret. The current trend is in favour of black-box models (i.e., neural network-based approaches to machine translation).

7 Conclusion

In conclusion, we have carefully distinguished between the task of translation (§1) and the responsibility of the translator (§3). In §1, we identified the two necessary and sufficient conditions (C1–C2) that must be satisfied in order for this task of translation to be successfully accomplished. We characterized natural languages in terms of their complexity, ambiguity, and diversity and distinguished between monolingual NLP tasks and machine translation tasks involving language pairs. In §2, we introduced the two horns (H1–H2) of the computational dilemma, argued that human translators do not embrace H1 indefinitely, and conceded that it remains unclear how human translators are nonetheless able to overcome the computational dilemma despite their memory constraints. In §3, we argued that the responsibility of the translator involves more than the merely successful performance of the task of translation: it equally involves a requisite level of natural language understanding. In §4, we distinguished between the metaphysics of understanding and the epistemology of understanding, related the former (concerned with what might hold in principle) to the latter (concerned with what might hold in practice) and argued that our gold standard for a metaphysical account of understanding remains either internalism, referentialism, some combination of both, or some alternative theory of meaning that is neither internalist, referentialist, nor pragmatist. In §5, we provided a survey of machine translation methods, including RBMT (§5.1), SMT (§5.2), NMT (§5.3), and foundation models (§5.4). In §6, we introduced Bar-Hillel's hypothesis about the impossibility of Fahqmt, relied on an example from Bender and Koller (2020), and defended the claim that none of the methods of machine translation ultimately succeed in overturning this hypothesis. Gaps between the linguistic form of training data and linguistic meaning remain, human-like understanding remains out of reach for state-of-the-art machine translation systems, Furthermore, human intervention (pre- and post-editing) is still part and parcel of the machine translation process, although it may equally be conceded that improvements in machine translation could make these tasks easier than they currently are. Once we drop the ideal of full automation, start thinking in terms of degrees of automation, and carefully consider how we might test for, track, and study the degree to which machine translation programs understand natural language (as opposed to merely processing it), we may end up having a far more fruitful discussion about the prospects and limitations of machine translation.

Curmudgeon Corner Curmudgeon Corner is a short opinionated column on trends in technology, arts, science and society, commenting on issues of concern to the research community and wider society. Whilst the drive for super-human intelligence promotes potential benefits to wider society, it also raises deep concerns of existential risk, thereby highlighting the need for an ongoing conversation between technology and society. At the core of Curmudgeon concern is the question: What is it to be human in the age of the AI machine? -Editor.

Data availability Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Azevedo Frederico AC et al (2009) Equal numbers of neuronal & non-neuronal cells make the human brain an isometrically scaled-up primate brain. J Comparat Neurol 513(5):532–541

Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align & translate. In: arXiv preprint arXiv: 1409.0473

Bar-Hillel Y (1960) The present status of automatic translation of languages. Adv Comput 1:91–163

Barreiro A et al (2011) OpenLogos machine translation: philosophy, model, resources & customization. Mach Transl 25:107–126

Bayes T (1763) An essay towards solving a problem in the doctrine of chances. Philos Trans R Soc Lond 53:370–418

Bender Emily M, Gebru T, et al (2021) On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In: Proceedings of the 2021 ACM conference on fairness, accountability, & transparency, pp 610–623

Bender Emily M, Koller A (2020) Climbing towards NLU: On meaning, form, & understanding in the age of data. In: Proceedings of the 58th annual meeting of the association for computational linguistics, pp 5185–5198

Bommasani R et al (2021) On the opportunities & risks of foundation models. In: arXiv preprint arXiv:2108.07258

Broca P (1861) New observation of aphemia caused by a lesion of the posterior half of the second & third left frontal convolutions. Bull Soc Anat (Paris) 6:398–407

Brown P et al (1988) A statistical approach to language translation. In: Proceedings of the 12th Conference on Computational Linguistics. COLING'88. Budapest, Hungry: Association for Computational Linguistics, pp 71–76. ISBN: 963 8431 56 3. https://doi.org/10.3115/991635.991651

Brown Peter F et al (1990) A statistical approach to machine translation. Comput Linguist 16(2):79–85

Brown T et al (2020) Language models are few-shot learners. Adv Neural Inf Process Syst 33:1877–1901

Chomsky N, Roberts I, Watumull J (2023) The False Promise of Chat-GPT. In: The New York Times. https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html (visited on 03 Aug 2017)

Council, National Research (1966) Language & Machines: Computers in Translation & Linguistics. https://doi.org/10.17226/9547

Devlin J et al (2019) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis, Minnesota: Association for Computational Linguistics, pp 4171-4186. https://doi.org/10.18653/v1/N19-1423. https://aclanthology.org/N19-1423

Duff Melissa C, Brown-Schmidt S (2012) The hippocampus & the flexible use & processing of language. Front Hum Neurosci 6:1–11

Elbourne P (2011) Meaning: A slim guide to semantics. Oxford University Press, Oxford

Gair James W (2006) Charles Francis Hockett, 1916-2000: A Biographical Memoir. National Academy of Sciences

Halliday Michael Alexander K (2003) On Language & Linguistics. Ed. by Jonathan Webster. A & C Black, London

Hockett Charles F (1959) Animal "languages'' & human language. Hum Biol 31(1):32–39

Hockett Charles F (1967) A Course in Modern Linguistics. Macmillan, New York

Kalchbrenner N, Blunsom P (2013) Recurrent continuous translation models. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Seattle, Washington, USA: Association for Computational Linguistics, pp 1700–1709. https://aclanthology.org/D13-1176. Accessed 14 April 2023.

Klebanov Beata B, Flor M (2013) Associative texture is lost in translation. In: Proceedings of the Workshop on Discourse in Machine Translation, pp 27–32

Koehn P (2009) Statistical Machine Translation. Cambridge University Press, Cambridge

Radford A et al (2021) Learning transferable visual models from natural language supervision. In: International conference on machine learning. PMLR, pp 8748–8763

Reifler E (1952) Mechanical translation with a pre-editor, & writing for MT. Proceedings of the Conference on Mechanical Translation

Rosenblatt F (1957) The perceptron, a perceiving and recognizing automaton (Project PARA). Cornell Aeronautical Laboratory

Scott B (2003) The logos model: An historical perspective. Mach Transl 18(1):1–72

Scott B (2018) Translation, Brains & the Computer: A Neurolinguistic Solution to Ambiguity and Complexity in Machine Translation, vol 2. Springer, New York

Sundararajan M, Taly A, Yan Q (2017) Axiomatic attribution for deep networks. In: International Conference on Machine Learning. PMLR, pp 3319–3328

Sutskever I, Vinyals O, Le Quoc V (2014) Sequence to sequence learning with neural networks. In: Proceedings of the 27th International Conference on Neural Information Processing Systems. MIT Press, pp 3104–3112

Tenney I, Das D, Pavlick E (2019) BERT rediscovers the classical NLP pipeline. In: arXiv preprint arXiv:1905.05950

Thrun S (1998) Lifelong Learning Algorithms. Learning to learn 8:181–209

Turing Alan M (1950) Computing Machinery & Intelligence. Mind 59(236):433

Vanmassenhove E, Hardmeier C, Way A (2019) Getting gender right in neural machine translation. In: arXiv preprint arXiv:1909.05088

Vanmassenhove E, Shterionov D, Way A (2019) Lost in translation: Loss & decay of linguistic richness in machine translation". In: arXiv preprint arXiv:1906.12068

Vaswani A et al (2017) Attention is all you need. Adv Neural Inform Process Syst 30

Vauquois B (1968) A survey of formal grammars & algorithms for recognition & transformation in mechanical translation. IFIP Congress 68:1114–1122

Wernicke C (1970) The aphasic symptom-complex: a psychological study on an anatomical basis. Arch Neurol 22(3):280–282

Wittgenstein L (1953) Philosophical Investigations. Wiley-Blackwell, New York