

Meeting them halfway: Altering language conventions to facilitate human-robot interaction

Lize Alberts 

Department of Philosophy, Stellenbosch University, South Africa
E-mail: alberts3217@gmail.com

Abstract

This article considers the remaining hindrances for natural language processing technologies in achieving open and natural (human-like) interaction between humans and computers. Although artificially intelligent (AI) systems have been making great strides in this field, particularly with the development of deep learning architectures that carry surface-level statistical methods to greater levels of sophistication, these systems are yet incapable of deep semantic analysis, reliable translation, and generating rich answers to open-ended questions. I consider how the process may be facilitated from our side, first, by altering some of our existing language conventions (which may occur naturally) if we are to proceed with statistical approaches, and secondly, by considering possibilities in using a formalised artificial language as an auxiliary medium, as it may avoid many of the inherent ambiguities and irregularities that make natural language difficult to process using rule-based methods. As current systems have been predominantly English-based, I argue that a formal auxiliary language would not only be a simpler and more reliable medium for computer processing, but may also offer a more neutral, easy-to-learn lingua franca for uniting people from different linguistic backgrounds with none necessarily having the upper hand.

Keywords: natural language processing, artificial language, symbolic AI, computational linguistics, international auxiliary language

1. Introduction

Ever since the idea of artificially intelligent (AI) agents emerged, researchers, philosophers, and science-fiction novelists have been concerned with the profound political, economic, and ethical implications it may hold for human life. As various task-specific forms of AI technologies are becoming increasingly prevalent, we may consider some other, perhaps less glamorous, areas that may be affected, one being our language usage. In this article, I explore how the increased use of communicative AI technologies may lead to changes in human language conventions. Based on current limitations of the statistical approaches to natural language processing (NLP), I predict some ways in which we may be inclined to naturally alter our current language usage so that it may be processed more effectively. Although this may facilitate the process, there may be inherent limitations to dealing with language on merely a

statistical level, especially when it comes to common-sense reasoning and open conversation. In various narrow applications of NLP, statistical approaches have proven very successful by extracting patterns from large sets of sample data. This has been enabling computers to deal with the complexity of natural languages without requiring extensive lists of explicit grammatical rules and exceptions. Some drawbacks are that solutions drawn from finite datasets are often superficial and so task-specific that they cannot be transferred to other domains, and the “rules” or patterns computers gauge from their given sample data are not necessarily the ones we would like them to. However, thorough rule-based approaches have proven too time-consuming and unreliable for processing natural language, and would be more effective given an inherently rational, regular language with definite grammatical rules. This is something that a formalised artificial language could offer, as existing examples such as Ido and Lojban already suggest. Therefore, given our goal, I also critically consider the possibility of such a system as an auxiliary medium for language processing.

My investigation consists of four main parts: first, I offer some context regarding the current trajectory of development in the field of language processing, and what we have been aiming towards. Secondly, I investigate the current strengths and limitations of NLP systems, given all the inherent, messy aspects of natural language that problematise its formalisation. Thirdly, based on these insights, I consider how our language conventions might be affected by increased interaction with such software. Here, I also consider: (i) how popular forms of communication technologies have already been affecting the communicative behaviours of its users, and (ii) how the conventions of a language (particularly English) naturally tend to be altered in communication with non-native speakers. As the latter are unable to engage with the language on the same intuitive level as native speakers, I draw parallels between them and NLP software. Finally, I consider the possible benefits (and drawbacks) of adopting a formalised universal auxiliary language, not only for communication between people and computers, but between people as well. Here, I also critically compare the “hits and misses” of some existing artificial languages, namely Esperanto, Ido, and Lojban. In the final section, I offer a few closing remarks on the scope of my argument, and conclusions regarding future possibilities in human-robot social interaction.

2. Background

Ever since the discipline began more than 70 years ago, one of the major struggles in computer science has been to make computers *literate*, that is, capable of interacting with us in natural language (Bose 2004: 1, Hartshorne 2011: 44, Waldrop 1984: 372). As envisioned in fictional talking robots such as Ash in *Alien*, HAL in *2001: A Space Odyssey*, and the replicants in *Blade Runner*, the ultimate aim for researchers in the field of NLP has been to design machines that are able to interpret and use our ways of speaking so naturally that one could communicate with them as easily (and openly) as with another person (Bose 2004: 2). The early pioneer of AI research, Alan Turing, predicted that something like this would already be actualised by the turn of the 20th century:

I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9 , to make them play the imitation game^[1] so well that an average interrogator will not have more than a 70 per cent chance of making the right identification after five minutes of questioning.
(Turing 1950: 442)

¹ Turing's “imitation game” (also known as the “Turing test”) is a test of a machine's ability to emulate human-like responses to the point of indistinguishability.

Alas, as yet, robots with *true* human-level conversational ability remain constricted to the realm of dreams and science fiction. It turns out natural language understanding is a lot more complex – and outright bizarre – than we could have anticipated. Tasks like using contextual cues to infer the correct sense of an ambiguous utterance seem simple to us, but prove difficult to program, as a lot of what goes on in our language conventions is not reducible to definite formal rules. Rather, these conventions are the battered and bastardised products of centuries of contingent evolution, with which we also tend to take a lot of pragmatic liberties. Therefore, the problems of successfully parsing and interpreting the complexity of English alone (having been the dominant focus of NLP technologies from the start²) remain largely unsolved, and “considerable effort is still needed to bring language technology to the desired level of a pervasive, ubiquitous and transparent technology” (Ananiadou, McNaught and Thompson 2012: 2). This is without even accounting for the multiple other living natural languages that would require an enormous amount of research just to match (Ananiadou et al. 2012: 1–2).

The multiplicity of languages poses another problem: the barriers caused by the cultural and linguistic diversity in Europe alone remain a pressing issue in our globalising world, not only on interpersonal levels, but also in the spheres of business, politics, and information. To overcome this, the EU institutions collectively spend around €1 billion per year on maintaining their multilingual policy through translation and interpretation services (Ananiadou et al. 2012: 1). Although foreign language acquisition would significantly aid the process, mastering even one additional natural language is difficult given the significant differences even between closely related languages. Translation was, in fact, the very first application of AI computers, a task that Leon Dostert – the researcher responsible for original translation techniques – believed would take only “five, perhaps three” years to master fluently and reliably (Hartshorne 2011: 44). For that too, we are still waiting.

This is not to say that we have not been making progress. In recent years, probabilistic models³ of language technologies have been gaining increasing levels of sophistication in their combination with machine learning and deep learning methods, supported by high-speed internet and cloud computing. As a result, the market for NLP software is growing⁴, and all the more users are starting to use speech-based AI assistants⁵ such as Siri (Apple), Google Assistant (Google), Cortana (Microsoft), and Alexa (Amazon), as well as voice-activated home speakers such as Google Home, Amazon Echo, and Apple Homepod. Not only are the latter able to answer queries, they can control other devices via the internet, paving the way for the proliferation of more smart devices forming part of the Internet of Things (IoT)⁶.

² The landscape of language technology has always been dominated by English resources: leading conferences and scientific journals for the period 2008–2010 reveals 971 publications on language technology for English, compared to 228 for Chinese, and 80 for Spanish. Automated translation systems that translate into English tend to be the most accurate (Ananiadou et al. 2012: 2).

³ These models generate output by calculating the probability of various possible outputs given collected data of how the language is used rather than relying on explicit prescriptive rules.

⁴ A recent Tractica report predicts an expected growth from US\$136 million in 2016 to US\$5.4 billion by 2025 (Madhavan 2018).

⁵ ComScore predicts that by 2020, 50% of all searches will be voice searches (Ramamurthy, Morya, Karthik, Vijay and Gupta 2017: 6).

⁶ IoT refers to a web-enabled network of devices that can transfer and act on data.

These speech-based technologies remain imperfect⁷, and the smart speaker market still has significant room left to grow⁸, yet it seems we are fast approaching a paradigm wherein all the more user interfaces and service operators take the form of communicative AI software (Bose 2004: 1, Bianzino 2017). At 2015's SOLID Conference in San Francisco, Andy Goodman, group director of Fjord, predicted that the future of user interface design will mostly be voice-controlled, haptic, and invisible (a concept he calls "Zero UI"), offering a more user-friendly and natural way of interacting with technology (Benson 2015). This is echoed by Accenture in their Technology Vision of 2017 in their chapter entitled "AI is the new UI" (Bianzino 2017).

The obvious limitation of statistical models, however, is that they are corpus-based: the quality of the output largely depends on the amount and quality of the available data. This means they are likely to fail in the case of languages that have a smaller body of training samples or sentences with complex or less common structures (Ananiadou et al. 2012: 2). Moreover, these processes mainly function on a surface level in that they treat languages as patterns, perhaps with some knowledge of language-specific grammar rules. As such, they are yet incapable of symbolic abstraction and symbol manipulation⁹, and thereby deeper levels of semantic analysis, something which many¹⁰ believe is required for next-generation systems¹¹. Despite the recent hype surrounding neural network models following their remarkable successes in various subfields of language processing, particularly that of categorisation, sceptics warn that we should not let this lead us to expect too much from systems that are designed only to deal with very specific problems (Marcus 2018: 20, Nield 2019). When it comes to dealing with more open-ended problems in the real world, current models reach their limitations.

Given the trajectory of our increasing interaction with AI systems, the question arises as to how we may reach the ultimate goal of enabling this interaction to proceed as freely and naturally as that between people, and hopefully not just for English speakers. My aim is to investigate the remaining gap between natural language conventions and machine learning capabilities, and consider how we may perhaps *meet them halfway*, not merely by teaching computers how to interpret our use of language, but also by adjusting our usage to suit their modes of interpretation. For this, I first look at some of the major approaches in NLP, drawing mainly from Russell and Norvig's (2010) "Artificial Intelligence: A modern approach", the leading textbook in the field of AI.

⁷ They still have an error rate of roughly 5% for processing simple commands in natural language (Boyd 2018).

⁸ 75% of US homes are predicted to have at least one smart speaker by the end of 2020 (Boyd 2018).

⁹ That is, manipulating symbols (and, consequently, the abstractions they represent) according to logical rules.

¹⁰ See, for instance, Ananiadou et al. (2012: 2); Marcus (2018); Young, Hazarik, Poria and Cambria (2018); and Nield (2019).

¹¹ These systems may include more sophisticated (human-like) AI companions or teachers.

3. Natural language processing

3.1 Language models

Given the free and contingent development of natural language conventions, these languages typically pick up multiple ambiguous¹², superfluous¹³, and irregular¹⁴ features. Apart from the typological and phonetic complexities of English that make it particularly difficult to master (especially as an additional language) – not to mention the arbitrariness of its rules and their exceptions – various common semantic and pragmatic ambiguities often require us to depend on our intuition or contextual inferences to interpret utterances. Despite our attempts, misunderstanding is not uncommon between us, and more so for AI parsers that lack much of these interpretive abilities. In the first part of this section, I briefly discuss how computers currently attempt to tackle NLP despite the messiness of natural language, particularly in terms of how language models are created, how these models are used to parse language data, how interpretations are disambiguated, and how the field is being transformed by the development of neural networks and deep learning.

The predominant NLP approach relies on the use of statistical language models¹⁵ (Russell and Norvig 2010: 860). The most basic of these is the n -gram model, a model that merely predicts the probability of a given sequence of units¹⁶ of length n . However, as the training data could only offer an estimation of the true probability distribution¹⁷, models should also account for the possibility of texts they have not seen before, and they should not claim it to be impossible. Therefore, sequences that have a count of zero are given a small, nonzero possibility, and, consequently, the counts of other sequences are slightly lowered so that the probability still sums to 1. This process is called “smoothing” (Russell and Norvig 2010: 862–863).

N -grams are often used for limited-scope language processing tasks such as language identification, spelling or grammar correction, named-entity recognition, and text classification. The biggest issue for these models, however, is data sparsity: “[in a language] with a vocabulary of, say, 10^5 words, there are 10^{15} trigram possibilities to estimate, and so a corpus of even a trillion words will not be able to supply reliable estimates for all of them” (Russell and Norvig 2010: 888). A more sophisticated language model addresses this problem by including notions of lexical and syntactic categories which are then combined into trees representing the phrase structure of sentences. A popular example is the probabilistic context-free grammar¹⁸ (PCFG) model which uses treebanks to determine the likelihood of parses via machine learning¹⁹. Again, this is based merely on probability, given the lack of rigid grammatical rules in natural languages: “We are

¹² This could either be due to lexical elements, where individual words may be interpreted in different ways, or structural elements, where there is more than one way of analysing the underlying structure of a sentence.

¹³ This refers to aspects such as wordiness, pleonasm, and repetition that add no essential information to an expressed proposition (see Wit and Gillette 1999 for further reading).

¹⁴ Although some language change may be motivated by the correction of irregularities or the reduction of morphological complexity, much of it is not – see Bryson (1990) and Hickey (2010) for further reading.

¹⁵ These are models that are used to predict the probability distribution of expressions in a language based on given corpora.

¹⁶ These could be characters, syllables, words, etc. If it deals with one unit at a time, it is called a “unigram”, “bigram” for a sequence of two items, “trigram” for a sequence of three, and so on. The larger the n , the better understanding the model would have of context (Russell and Norvig 2010: 861).

¹⁷ This is not only because they are limited, but also because natural language conventions vary and change.

¹⁸ “Grammar” here is defined as a collection of rules that define a language as a set of allowable strings of words.

¹⁹ With machine learning, a computer system makes predictions and inferences based on algorithms and statistical models (created from its training data) in order to perform tasks without explicit instructions.

unlikely ever to devise a complete grammar for English, if only because no two persons would agree entirely on what constitutes valid English” (Russell and Norvig 2010: 890).

A simple PCFG may determine the probability of each individual word in its lexicon (list of allowable words) of belonging to particular lexical categories²⁰, as well as the probability of each lexical category constituting a particular syntactic category²¹. Based on these, a PCFG may be able to generate grammatical sentences relatively accurately. However, other factors have to be considered as well: for example, the form of individual words may differ based on their relative placement in sentences. To illustrate, “I like her and she likes me” is grammatical, but “Me likes she and her likes I” is not. Similar to the limitation of *n*-grams, it may also undergenerate, that is, not recognise grammatically correct constructions it has not encountered before (Russell and Norvig 2010: 890–892).

Another limitation for PCFGs pertains to (syntactic) context sensitivity: making judgments based on lexical and semantic categories, and not the relation between the meanings of words themselves²² (Russell and Norvig 2010: 897). To address this problem, a PCFG can be lexicalised to determine the likelihood for words in particular relations. As this would not be feasible for probabilities to depend on every word in a sentence, grammar is *augmented* so that only the head of each phrase is analysed for probability in relation to one another. Of course, this means that things like nonsensical adjectives would still not be caught by these models. Another issue is that, once again, given the vastness of vocabularies, a corpus would not be able to account for each possible relation, and most of the estimates would have to come from smoothing (Russell and Norvig 2010: 897–898).

Grammars can be augmented further in terms of semantics. One approach is using models based on predicate logic, which indicates the particular relation between syntactic units in terms of rules written as logical statements. The semantic analysis system then draws conclusions for its interpretation based on the meaning representation of an expression and its match in a knowledge base (Bose 2004: 6). Meaning representations are formed by formalising expressions in terms of objects (the heads of noun phrases) and relations (properties such as *red* or *beautiful*, or more complex relations such as *smaller than* or *underneath*, or functions such as *mother of*) (Russell and Norvig 2010: 288). For example, the sentence “HAL needs batteries” may have something like *Needs(HAL, Batteries)* as its semantic interpretation, with *Needs* being a particular relation between the subject *HAL* and the object *Batteries*. The verb phrase “needs batteries” is a description that serves as a function which may or may not apply to a particular entity – this could either be drawn directly from a knowledge base or inferred from other similar representations (Bose 2004: 6). In lambda notation, the relation can be expressed as $\lambda x \text{ Needs}(x, \text{ Batteries})$, with λ indicating that the variable *x* is bound in the expression. A rule would then be added that a *noun phrase* with the semantic role of *object*, followed by a *verb phrase* with the semantic role of *predicate*, yields a sentence of which the semantics is the result of applying *predicate* to *object*. Russell and Norvig (2010: 902) represent this as:

$$S(\text{pred}(\text{obj})) \rightarrow \text{NP}(\text{obj}) \text{ VP}(\text{pred})$$

²⁰ For instance, a pronoun can have a probability of 10% to be *me*, of 10% to be *I*, of 3% to be *you*, and so on.

²¹ For instance, a noun phrase can have a 30% chance of being constituted by a pronoun, 25% for a noun and article, etc.

²² For example, it may see no problem with “I want a glass of angry telephone” as, structurally, the sentence is correct.

Therefore, the rule would make the semantic interpretation of “HAL needs batteries” as $(\lambda x \text{Needs}(x, \text{Batteries}))(HAL)$ which states that *HAL* replaces the variable x in the function. To add notions of time, Russell and Norvig (2010: 903) suggest representing different tenses in the form of event calculus notation²³:

John loves Mary: $E_1 \in \text{Loves}(\text{John}, \text{Mary}) \wedge \text{During}(\text{Now}, \text{Extent}(E_1))$
 John loved Mary: $E_2 \in \text{Loves}(\text{John}, \text{Mary}) \wedge \text{After}(\text{Now}, \text{Extent}(E_2))$

These serve to distinguish between two tenses: the simple present and the simple past. The authors explain that these can then be turned into lexical rules (Russell and Norvig 2010: 903), for instance:

$\text{Verb}(\lambda y \lambda x e \in \text{Loves}(x, y) \wedge \text{During}(\text{Now}, e)) \rightarrow \text{loves}$
 $\text{Verb}(\lambda y \lambda x e \in \text{Loves}(x, y) \wedge \text{After}(\text{Now}, e)) \rightarrow \text{loved}$

In the *HAL* example, the verb *needs* is in the present perfect tense, therefore the rule which would apply may also be something like $\text{Verb}(\lambda y \lambda x e \in \text{Needs}(x, y) \wedge \text{During}(\text{Now}, e)) \rightarrow \text{needs}$. However, this does not nearly account for all the tense distinctions in English, and further rules would be needed to discern between other forms of the verb, for instance, between singular and plural forms. As noted before, grammatical rules in natural languages are also bound to have exceptions which would have to be solved using probability based on statistics (Russell and Norvig 2010: 903). Also, given that the rules for each natural language differ, the semantic interpretation of another language would require the formulation of a whole new set of rules and exceptions.

The examples above are still relatively straightforward to interpret, however, this is not often the case in natural language expressions. The biggest obstacle is the various forms of lexical and syntactic ambiguity found in natural languages which could cause a computer to end up with multiple possible parses or semantic interpretations of a given sentence (Russell and Norvig 2010: 902–905). For instance, Russell and Norvig use the example “Every agent feels a breeze” which has only one syntactic parse but two semantic interpretations: it could either be understood as *For every agent there exists a breeze* or *There exists a breeze that every agent feels* (Russell and Norvig 2010: 903). Syntactic ambiguity may also lead to semantic ambiguity given the fact that there is more than one parse for a given expression²⁴ (Russell and Norvig 2010: 905). Such ambiguities are impossible to resolve merely with rules – a computer may need to appeal to other factors like knowledge about particular words²⁵, contextual clues, or real-world knowledge (Hutchins and Somers 1992: 91). Although we tend to function quite well using the latter two, for statistical and/or rule-based computers it is more difficult. In terms of drawing from contextual knowledge, Hutchins and Somers (1992: 92) contend that the difficulty for NLP systems is that, once again, there are no definite rules for where the appropriate information is to be found for each individual case. Even if it is possible to store

²³ \in is used to indicate *there exists*, and \wedge is used to indicate a logical conjunction.

²⁴ For instance, “The woman saw a man with the telescope” could be parsed either so that “with the telescope” describes the *verb*, or the *object* (“the man”).

²⁵ That is, context-independent information about words and how they combine with others. This could entail providing parsers with information about co-occurrence restrictions; for example, indicating the types of complements that are *expected* to go with particular verbs or the types of nouns that generally fill particular syntactic roles (Hutchins and Somers 1992: 91–92). On a more general level, this could also be handled in terms of verb valency or case grammar.

previous knowledge derived from the text²⁶, it would still be unclear which information may count as useful, and “it would clearly be impractical to extract and store every fact that could be inferred from every sentence of a given text, just in case it was needed to disambiguate something” (Hutchins and Somers 1992: 92–93). Failing that, real-world knowledge can also be useful to discern which reading seems realistically most probable²⁷ (Hutchins and Somers 1992: 93, Russell and Norvig 2010: 906). However, Hutchins and Somers (1992: 93) also maintain that it is impractical to program and incorporate all real-world knowledge that may potentially help to disambiguate such statements, even in narrow applications, and even with more advanced technology. Another factor to consider is the likelihood that the speaker intends to communicate the particular fact to the hearer. For example, Russell and Norvig (2010: 906) explain that the real-world knowledge approach may assign a higher probability that “I am not a crook”, when uttered by a politician, refers to a hooked shepherd’s staff than a criminal, as the former is factually less probable, although the latter is more likely to be the intended sense.

Failing these, Hutchins and Somers (1992: 94) suggest that a system may make use of some strategies that a human interpreter might use in the same circumstances. One such option is simply asking the author or speaker directly; another is what they call the “best guess” strategy, which is to determine the most likely interpretation based on whichever sentence structures are most common, regardless of the specific words involved (Hutchins and Somers 1992: 94). Of course, although guesses may be well motivated, this is not exactly a reliable method. Especially since most utterances in natural language are actually highly ambiguous, even if it may not be apparent to native speakers, and a system with a large grammar and lexicon may find thousands of interpretations for an ordinary sentence (Russell and Norvig 2010: 906).

Given how time-consuming, and often inaccurate, the process of constructing hand-crafted rules for natural language usage turned out to be, recent NLP research has increasingly been focusing on the use of word embeddings, neural network²⁸ language models that automatically create distributed feature vector representations²⁹ for words by extracting their functional characteristics from large sets of word sequences (Young et al. 2018: 55–56). The aim is to learn the contexts in which each word may be used by capturing the features of its neighbours³⁰. The main advantage of these feature vectors is that they can efficiently capture the similarity between words³¹, which helps the network to predict the probability distribution over the next word in a sequence, as well as which words can replace each other in similar contexts. Another benefit is that they reduce the impact of the “curse of dimensionality”³² of former statistical models: rather than requiring at least one example for each relevant combination of the input variables, the distributed representation approach allows the model to generalise better to

²⁶ In recent years, Long Short-Term Memory (LSTM) neural networks have started to address this problem. Normal LSTM networks are able to store about 30 time steps, but not a whole page, let alone a whole book (Goyal, Pandey and Jain 2018: 152).

²⁷ For instance, in “The woman saw a planet with the telescope”, it is more likely for a woman to look at a planet with a telescope than for a planet to have a telescope, thus the former reading is probably the intended one.

²⁸ A neural network is an interconnected system of simple processing units, or “nodes”, loosely based on the way neurons interact in the animal brain. The relative connection strengths, or “weights”, between nodes are obtained and modified by a process of learning from a set of training patterns (Gurney 2004: 13–16).

²⁹ Based on extracted functional features, each word is associated with a continuous-valued vector representation in a low-dimensional space where every dimension corresponds to a different feature (Bengio 2008: 3881).

³⁰ For instance, the CBOW model computes the conditional probability of a target word given the context words surrounding it across a window of size k (Young et al. 2018: 57).

³¹ This is as functionally similar words will, at least in some directions, be closer to each other (Bengio 2008: 3881).

³² That is, the exponential increase in the number of required training examples as the number of input variables increases.

sequences that are not in the training set but have similar features (Bengio 2008: 3881). This has been further improved on by the use of deep learning³³ methods which enables multi-level automatic feature representation learning (Young et al. 2018: 55).

Despite these relative advantages, a limitation of individual word embeddings is their inability to represent phrases of which the meanings are not reducible to the combined meanings of their parts³⁴. Another is introduced when embeddings are learnt based only on a small window of surrounding words, and so semantically-similar words that express opposing sentiments may be clustered together³⁵, which is particularly problematic for tasks that require sentiment analysis. Models that assign a global embedding to each word also suffer problems like being unable to account for polysemy, although some deeper networks have started providing different representations for varying senses of the same word (Young et al. 2018: 59). Moreover, Young et al. (2018: 59) point out that a general caveat for word embeddings is that they are highly task-specific, and training them from scratch for a new application requires a lot of time and resources. In recent years, discussions have emerged on the relevance of distributional feature vectors in the long run; a small, but growing consensus³⁶ in the AI community suggests that adequate representations of words and concepts cannot be inferred from distributional semantics alone.

Considering deep learning in general, Marcus (2018) discusses 10 challenges this approach currently faces: that it is data-hungry, has superficial solutions³⁷ with a limited ability for transfer, has no natural way to deal with hierarchal structure³⁸, struggles with open-ended inference, is not sufficiently transparent³⁹, is not well integrated with prior (real-world) knowledge, cannot inherently distinguish causation from correlation, presumes a largely stable world, cannot be fully trusted⁴⁰, and is difficult to engineer with⁴¹. He considers many of these extensions of the fundamental problem of contemporary (mainly supervised⁴²) deep learning systems: that they do well on challenges closely resembling their training data but less well on more open-ended cases or those on the periphery which often occur in the real world (Marcus 2018: 16). By over-hyping the successes of these models, he raises the concern that “the field of AI could get trapped in a local minimum [...] focusing too much on the detailed exploration of a particular class of accessible but limited models that are geared around capturing low-hanging fruit – potentially neglecting riskier excursions that might ultimately lead to a more robust path” (Marcus 2018: 18). Instead, he suggests that the above challenges could be addressed by integrating the use of

³³ Whereas shallow neural networks only have one hidden layer (that is, a layer of neurons between the input and output layer), deep networks have multiple. This allows for more complex correlations to be extracted from the training data.

³⁴ For instance, idioms like “hot potato” or named entities like “Boston Globe” (Young et al. 2018: 59).

³⁵ For example, words like “good” and “bad” can share almost the same embedding (Young et al. 2018: 59).

³⁶ See, for example, Kiela, Bulat, Vero and Clark (2016), Gauthier and Mordatch (2016), and Lucy and Gauthier (2017).

³⁷ Recent experiments have shown that the performance of various deep networks trained on a question-answering task dropped precipitously with the mere insertion of distraction sentences (Marcus 2018: 8–9).

³⁸ That is, syntactic relations between main clauses and embedded clauses in a sentence (Marcus 2018: 9).

³⁹ Rather than using parameters that we can clearly interpret and control, the features extracted by hidden layers are opaque and less straight-forward, which can lead to strange biases in algorithms (Marcus 2018: 10–11).

⁴⁰ Given how deep learning systems base their inferences on features they pick up on in training data, rather than explicit definitions, they can be easily fooled (e.g. mistaking yellow and black stripes for school buses) (Marcus 2018: 13–14).

⁴¹ Although machine learning is effective in limited circumstances, it will not necessarily work in others as it yet lacks “the incrementality, transparency and debuggability of classical programming” (Marcus 2018: 14).

⁴² *Supervised* machine learning systems learn from labelled datasets, i.e. sets of example input-output pairs, as opposed to *unsupervised* systems that find patterns in previously unseen (unlabelled) data. Whereas some word embeddings can be used in unsupervised settings, supervised learning is the most popular practice in recent deep learning NLP research, and unsupervised schemes are still in their developing phase (Young et al. 2018: 73).

symbolic systems: not just making informed guesses based on finite training examples, which is still useful for some applications, but learning how to represent abstractions and how they can be logically manipulated⁴³ (Marcus 2018: 20). Likewise, Young et al. (2018: 73) predict that coupling sub-symbolic AI (that is, using connectionist, deep-learning approaches) and symbolic AI (that is, with explicit symbolic programming) “will be key for stepping forward in the path from NLP to natural language understanding”.

In this subsection, I investigated the use of different statistical language models for a variety of NLP tasks, as well as their respective limitations. Another important NLP application is machine translation wherein fine interpretative distinctions can especially make a considerable difference. In the next subsection, I investigate some major current approaches in this field and their respective shortcomings.

3.2 Machine translation

Machine translation (MT) can be defined as the automatic translation of text from one natural language (the source language, SL) into another (the target language, TL). It was one of the first tasks of early computers but has only recently been gaining widespread usage (Russell and Norvig 2010: 907). In this subsection, I consider various MT approaches, focusing on four major types: the rule-based approach, which includes the direct, interlingual, and transfer-based methods; the corpus-based approach, which includes the statistical and example-based methods; the hybrid approach, which is a combination of the first two approaches; and finally, the recent development of neural MT, which includes the use of deep learning methods.

The rule-based (also called the “knowledge-based” or “classical”) approach generates output sentences in the TL based on linguistic (morphological, syntactic, and semantic) information about the respective languages, and was manually developed over time by human experts (Karami 2014: 1). This process consists of three stages: analysis, transfer, and generation, and the three main strategies that fall under this category differ in terms of the relative sizes of each of the three components (Hutchins and Somers 1992: 72). The first and most basic approach is the direct (word-for-word) method which has the least thorough analysis, and generates output through a direct translation of each element in the source text, largely irrespective of syntactic structure or true semantic equivalence (Hutchins and Somers 1992: 72). It consists of a morphological analysis of the SL sentence which is then translated by finding the corresponding words in bilingual dictionaries. This may be followed by some local rule-based reordering of elements such as noun complements or verb particles, and then the output is generated in the TL (Hutchins and Somers 1992: 72). Naturally, this approach has severe limitations which led to the development of more indirect MT models that use added intermediate steps. The first of these was the interlingual method, which abstracts away from the particular characteristics of the SL text, and focuses purely on semantic intent⁴⁴ (Dorr, Hovy and Levin 2004: 375).

This model can be divided into two components, namely an analysis of the SL text into an abstract, language-independent representation of meaning, and the generation of the semantic equivalent of this text into the TL (Hutchins and Somers 1992: 74). It makes use of a collection of representation symbols that (either independently or collectively) denote particular aspects

⁴³ This more closely resembles the PCFG approach, as well as that of programming languages where lines of code represent certain operations over variables that a computer can reliably interpret and execute (Marcus 2018: 20).

⁴⁴ As mentioned, this is ideally what future systems would have to be capable of for us to converse with them more openly.

of meaning by means of formal notation. Each element in the lexicons of the respective languages is directly or indirectly associated with one or more of these symbols (Dorr et al. 2004: 377). Initially, the aim was to develop a truly universal interlingual representation that can serve as an intermediary between all natural languages, which would enable multilingual translation within the same system (Hutchins and Somers 1992: 75). However, given the complexities involved, this aim came to be regarded as too ambitious, and few interlinguas are more than demonstration prototypes (Dorr et al. 2004: 375).

The second indirect rule-based approach is the transfer-based method in which the software parses and converts the SL text into an abstract intermediate (language-specific) SL representation which is then converted into an abstract (language-specific) TL representation, and then into a TL output (Hutchins and Somers 1992: 75). This process requires extensive lexicons with morphological, syntactic and semantic information, and complex sets of rules that are used to transfer the grammatical structure of the SL text into the TL (Karami 2014: 2). In comparison with the interlingual approach, this is a lot more complex, especially if more languages are added, as a third language would require four new transfer modules to be translatable into the other two, a fourth would require six more, and so on (Hutchins and Somers 1992: 75–76). The customisation cycle needed to reach the quality threshold may also be quite a costly and time-consuming process (Karami 2014: 2).

A more successful MT approach is the corpus-based (probabilistic) method, which is further divisible into the statistical method, based on the frequency of words and word combinations, and example-based method, based on the extraction and combination of phrases or text fragments, both of which rely on corpora of bilingual texts (Hutchins 2005: 197). Between these, the statistical method has become the dominant approach in MT. It uses computer algorithms that consider millions of possible ways of combining fragments of text to produce the most probable translation based on the analysis of both monolingual and bilingual corpora. Not only does this require a lot of processing power and extensive hardware configuration, it also requires *a lot* of data – a minimum of 2 million words are needed just for a specific domain in a single natural language (Karami 2014: 1). Typical drawbacks of models of this type include inconsistency and unpredictability, as the quality depends on what the model is able to guess based on the corpora (Karami 2014: 2).

In order to improve quality as well as cost and time efficiency, many rule-based MT developers have been combining their core technology with that of statistical models into what is referred to as “hybrid MT”. This approach takes the best of both models to compensate for what the other lacks: the good, out-of-domain quality and consistency of rule-based MT with the domain-specific, efficient, and cheap automaticity of the corpora-based method (Karami 2014: 2). The coupling of the two models is generally done in either a serial or a parallel pattern. In serial coupling, translations are performed using rule-based methods after which the output is edited using a statistical postprocessor based on bilingual corpora. Possible errors may be introduced by the postprocessor if, for instance, it omits some necessary elements or misinterprets the text (Xuan, Li and Tang 2012: 3018). In parallel coupling, data are weighted from all sources (statistic corpora and linguistic rules) and output is generated that satisfies both the most, using either module as a skeleton. However, this approach is difficult to use in practical applications as it is heavy on computational resources, and the vast number of output options to consider slows down the decision-making process (Xuan et al. 2012: 3019–3020).

A recent, major advancement in statistical MT is the development of neural network models. This advancement promises a better statistical command of synonymous words and a higher sensitivity to context. Whereas traditional statistical methods are based on a linear model⁴⁵, neural MT models are able to account for more complex context-specific relationships between features by using hidden layers of feature extraction that detect these patterns automatically (Koehn 2017: 7–10). This approach has emerged as the most promising in MT, and has shown superior performance on public benchmarks, yet it still has various challenges to overcome, most notably in out-of-domain⁴⁶ performance and dealing with conditions that differ significantly from training samples (Koehn 2017: 91). Other inherent weaknesses of neural models include slow training and inference speed, and the occasional failure to account for all elements in the input sentence (Wu et al. 2016: 1–2).

Google’s neural MT system (GNMT) has been attempting to address some of these issues, and has shown a promising 60% reduction in translation errors (in comparison to its phrase-based system) in a test involving the translation of 500 sampled sentences from Wikipedia and news websites. However, as anyone that has tested the Google Translate service with some multilingual language would know, it remains far from flawless. For instance, consider the simple Afrikaans sentence “My lief, jy kan nie spel nie” [‘My love, you can’t spell’]. If you input the parts on either side of the comma separately, the service⁴⁶ gets it correct, yet if you input the full sentence, Google translates it as “Love me, you can’t spell”⁴⁷. According to the SL, both interpretations should be the same. Strangely, if you change the sentence into the positive, “My lief, jy kan spel” [‘My love, you can spell’], Google⁴⁶ translates it as “Love me, you can play”, which is based on an incorrect translation of the word “spel” that can mean “game” but never “play”. If the same input is translated to Dutch, however, the result is still wrong even when the parts are separated: “My lief” [‘My love’] becomes “Hou van me”⁴⁶ [‘Like me’]. These examples show that the issues of inconsistency and unpredictability of statistical models still persist. It is also noteworthy that the results of various bilingual human translators, used as bases of comparison for Google’s test, were also not scored as reliably accurate owing to “possible ambiguities in the translations and also possibly non-calibrated raters and translators with a varying level of proficiency” (Wu et al. 2016: 19).

In this subsection, I compared four major types of MT systems in terms of their respective strengths and weaknesses. Although the rule-based approach would theoretically be most likely to offer grammatically correct sentences, the irregularity of natural language grammar and possible lack of equivalences between languages makes this largely unreliable, and makes corpus-based or hybrid approaches more (though still not completely) reliable. The same applies to all forms of NLP software: even the most sophisticated systems rely heavily on probabilistic reasoning to deal with natural language expressions, as attempts to make abstract representations of meaning based on rules are not only expensive and time-consuming, but also not fully reliable – the link between signifier and signified is not quite as straightforward as, for instance, in formal computer languages. Despite their increasing levels of intricacy and

⁴⁵ These models usually combine different, manually weighted components of an MT system (such as the language model, phrase translation model, the reordering model, etc.) and properties of the sentence (such as length) in a way that assumes them to be linearly separable (Koehn 2017: 7).

⁴⁶ In different domains (such as medicine, law, etc.), meaning is expressed differently. For out-of-domain performance, Koehn (2017: 91) found neural MT systems worse than statistical systems in almost all cases, and sometimes drastically so.

⁴⁷ <https://translate.google.com/> (Accessed 7 July 2019).

sophistication in limited applications, the inability to effectively discern the *meaning* of expressions and engage with them accordingly may be their greatest hindrance to achieving the broader scope of human-like conversational ability.

Another important limitation of NLP systems is the fact that they are still predominantly English-based⁴⁸, which means that non-native English speakers may not be able to use them with the same level of ease or success, or have access to similarly advanced technologies in their own languages soon (if ever). In response, Ananiadou et al. (2012: 2) maintain that building systems that are able to analyse the deeper semantic properties of language is the only way forward if we want to build reliable platforms that offer the same quality for non-English speaking users. However, given the particularities of each existing natural language, finding ways of achieving this for all of them seems like a tall order, especially since we are yet to succeed with even one of them. Perhaps more effective and reliable semantic interpretations could have been possible had we spoken a language more closely resembling a formal artificial language like Java – I explore this further in the final section. First, I consider how the increased use of NLP systems may affect our natural language conventions based on their current strengths and limitations as discussed above.

4. Modifying natural language

4.1 Technology and communication

Technology tends to have a profound effect on all aspects of human life, and language is no exception. In the first part of this section, I investigate how the increased use of digital communication technologies (like instant messaging and social networking) has been found to affect our communication behaviours. Secondly, I look at some examples of alterations to English brought on by non-native speakers as – being unfamiliar with complex sentence structures, and culture-specific habits and colloquialisms – these speakers typically rely more on simplified, unambiguous, literal formulations – simplifications I believe might benefit a computer parser as well. Drawing from these, I then make predictions for how increased interaction between humans and AI interlocutors in natural language (particularly English) may possibly affect our language conventions in future.

Instant digital communication can be described as operating on some middle ground between written language and the informality/contextual dependency of spoken language, which has been found to affect the micro communication behaviours and strategies of users⁴⁹ (Omar and Miah 2012: 13). According to Watt (2010: 144), the lack of contextual and nonverbal cues encourages users to provide more contextual information which enhances their pragmatic skills. Users have also been found to adapt their language use to suit the language environment, such as altering their “relational tone, personal language, sentence complexity, and message composition time depending on their target recipient” (Watt 2010: 144). These technologies have also led to new varieties of written language, such as netspeak, which deviates from the grammatical and syntactic rules of written English as a more informal, concise, abbreviated form that reads as if it were being spoken (Omar and Miah 2012: 9–15, Watt 2010: 141–143). The increased concision may also be attributed to constraints on screen size and character

⁴⁸ This is not only because English is one of the major world languages, but also because the dominant actors in the field of NLP are primarily privately-owned enterprises based in the US (Ananiadou et al. 2012: 2).

⁴⁹ The effects are especially profound on children as their skills are still developing (Watt 2010: 142).

allowance. Another effect of screen-based technologies, which particularly raises concern, is the decrease in face-to-face interaction (Omar and Miah 2012: 13–14). However, as these technologies keep evolving rapidly, their effects on us are likely to keep changing as well (Omar and Miah 2012: 15, Watt 2010: 147).

The influences discussed so far have mostly been limited to written language as this is the area in which digital communication platforms have had the most profound effect. They have also pertained particularly to interpersonal communication as most human-computer interaction (whether typed or spoken) has still been limited to closed questions and simple keyword-based searches. As future human-computer interaction is predicted to occur on a more natural, voice-operated basis, blurring the distinction between human-human and human-robot interaction, our spoken language behaviours are more likely to be affected (including, perhaps, our use of nonverbal cues such as gestures, facial expressions, and eye movement). It is difficult to make any strong predictions as it is unclear how sophisticated future systems may be and to what extent they will really be pervasive. However, based on my investigation into the current capabilities of NLP systems, it is possible to discern which areas are likely to be affected. To ground my argument, I first look at some examples of other simplified versions of natural language (particularly English) that have emerged as a result of non-native speakers struggling to deal with the complexities of the conventional language. I consider this to be an apt metaphor for how we may (at least initially) approach natural communication with computers.

4.2 Aiding comprehension for non-native speakers

As mentioned, my comparison between a non-native speaker of a language and an AI parser/interpreter is founded on the lack of contextual, culture-specific or colloquial conventions that enable native speakers to make intuitive inferences that may not be immediately evident from the statements themselves. As a result, these often have to be formulated more explicitly as these speakers generally tend to rely on more literal interpretations, smaller vocabularies, and a more basic comprehension of sentence structures. In this subsection, I investigate such strategies used by non-native English speakers to facilitate their mutual understanding (or between them and native speakers), from which I draw some that I believe may prove useful in the case of human-robot interaction as well.

As English is increasingly used by people from immensely varied language backgrounds, multiple new varieties emerge that are easier for non-native speakers to use and interpret. The overarching tendency is towards simplification, depending on lower lexical diversities⁵⁰ or forcing regularisation by analogy, although some complexities may also be added depending on what the speaker is used to (Bentz, Verkerk, Kiela, Hill and Buttery 2015: 18, Mauranen 2015: 37). The characteristic strategy is to enhance explicitness by adding redundant elements so as to enable effective communication despite possible errors (Mauranen 2015: 37–40).

These processes tend to occur automatically, although various attempts have been made to purposefully construct a simplified, *closed* version of English so as to standardise it. A popular example of such a version is Globish (**‘Global English’**), which is used internationally by both native and non-native English speakers to facilitate communication between them. It relies mostly on English words and phrases that are common throughout the English-speaking world, while

⁵⁰ This is done by relying on fewer word forms or more word frequencies.

minimising idiomatic phrases, figurative meanings, and ambiguities. Speakers are also encouraged to avoid metaphors, abbreviations, even humour – anything that might cause cross-cultural confusion. The concept was developed in 2003 by Jean-Paul Nerrière, a former IBM executive, who observed simplifications that enabled non-native English speakers to communicate with each other more successfully than with native speakers. Globish is based on a vocabulary of a mere 1500 words, involving a modular method for combining them, with an emphasis on concision, basic syntax, active verb tense, and the correct use of syllable stress (Clark and Gregor 2012: 24–25). Given its limited vocabulary, it relies on a larger number of words to communicate effectively without ambiguity or incomprehension⁵¹. Moreover, the use of gestures and facial expressions is encouraged to facilitate comprehension further (Nerrière 2003: 60).

These examples serve to show how, despite the inherent complexities of a natural language such as English, there are steps one can take to guard against miscommunication when talking to someone who lacks a native speaker’s ability to intuitively grasp the meaning of localised, ambiguous or overly complex expressions. Drawing from these insights, in what follows, I consider which strategies may aid the effective process of natural language expressions by statistical NLP systems, and how this might affect our language conventions in the future.

4.3 Future predictions

As noted before, the most effective strategy to prevent a non-native speaker (read: computer) from interpreting an ambiguous statement incorrectly is enhanced explicitness⁵² which may also include strategically added redundancy⁵³, and the avoidance of homonyms and polysemic or idiomatic expressions. As when communicating on an online platform, it would also help to provide necessary contextual information rather than omitting it (by, for instance, limiting the use of indexicals) given that inferring from context is still too intuitive and bizarre for NLP systems to do reliably.

One difference between a statistical language model and a second-language speaker is that the former would not necessarily benefit from *lexical* simplification, as the size of its vocabulary would probably exceed even that of a native speaker⁵⁴. Instead, what would be useful is *syntactic* simplification, and a reliance on the most frequently used grammatical patterns. The shorter and more basic a sentence structure is, the easier it is to parse and interpret, as not only do longer sentences tend to have higher levels of ambiguity, but more words also have to be analysed in terms of their relations to others (Siddharthan 2006: 99). For this reason, enhanced explicitness should only be employed where misinterpretation is a real threat in order to prevent making statements needlessly cluttered and complex. In short, the aim should be to convey the critical units of information and their relations as simply, clearly, and predictably as possible.

These are merely a few examples of possible language modifications that users of communicative AI systems may benefit from based on the current capabilities of NLP software.

⁵¹ For example, as the word “cunning” does not feature in the Globish vocabulary, the concept can be communicated through combining terms like “wise” and “hard to trust” (Nerrière 2003: 60).

⁵² For instance, using the previous example, rather saying, “There exists a breeze that every agent feels” if that is the intended sense.

⁵³ For example, “I want that glass *over there*”.

⁵⁴ However, given that NLP depends largely on probabilistic reasoning, limiting lexical diversity would be useful, as the more a particular word is used in a certain context, the more accurately the system would be able to predict its intended sense.

As has been the case with instant messaging, my prediction is that increased and prolonged use of such systems may also gradually alter our general communicative behaviours and language conventions. Drawing from the considerations above, this could entail enhanced explicitness, increased concision, and a greater reliance on basic grammatical structures and standardised word usage. Although the addition of some redundancies may aid interpretation, others may be removed as, ideally, one would seek the most efficient way of communicating key points of information. This could cause the use of some polite (yet superfluous) interpersonal communication habits, like greeting, thanking or apologising, to decrease as well. Given the predominance of English in the field, other effects may include a further increase in English acquisition, correlating perhaps with increased use of borrowed terms. If users become increasingly comfortable with verbal rather than written modes of communication⁵⁵, another possibility is that some of the aforementioned effects of screen-based communication technologies on written language, such as the use of netspeak, may be reversed.

Although some of the modifications discussed above may make natural language easier to process, some complex and ambiguous elements may still remain, and native speakers may find it difficult to always limit their habitual language usage in such ways. For a truly regular, easy-to-parse medium (and one that is easy to make symbolic representations of), one option is to rather communicate in an artificial language that is inherently formalised and less ambiguous – the topic of the next section.

5. Possibilities in artificial language

Unlike natural languages, artificial or constructed languages do not develop naturally through a community of speakers. Rather, they are created with specific aims according to definite grammatical, syntactic, and phonological rules. This not only makes the language easier to learn, but also does away with a lot of the ambiguity and irregularity found in natural languages. In this section, I critically discuss a few existing examples, namely, Ido, Esperanto, and Lojban, to consider what may be taken from the respective “hits and misses” of each. I then critically evaluate the possible benefits and drawbacks of standardising such a formalised, artificial auxiliary language.

5.1 Existing attempts

The most popular living example is Esperanto, an artificial language that Dr L.L. Zamenhof developed in 1887 with the aims of being exceptionally easy to learn, and to facilitate international communication and thereby global harmony. It has a root vocabulary of 917 words and 16 key rules of grammar, and was designed to be easily accessible and free from the complexities and irregularities of natural languages (Tellier 2013: 10–11). The root words are drawn from a combination of major European languages⁵⁶ which are used to express a variety of concepts through systematic derivation⁵⁷. This kind of structure gives the language an element of logical clarity as well as economy, as few words and rules are needed to express

⁵⁵ This is perhaps already reflected in the recent increase in the use of voice/audio notes on WhatsApp and Facebook. In fact, the latter is currently aiming to venture into hardware with its own voice-activated smart speaker, Portal (Luckhurst 2018).

⁵⁶ From the Romance group, these are mainly French and Latin; from the Germanic, mainly German, English, and Yiddish; and from the Slavic, mainly Russian and Polish, including a bit of ancient Greek (Gobbo 2017).

⁵⁷ For instance, the prefix *mal-* can be added to a word to form its opposite.

various shades of meaning (Dyer 1923: 91–93). Other useful features include its direct grapheme-phoneme (one letter, one sound) correspondence, and the transparency of its grammatical elements⁵⁸ (Tellier 2013: 10–11). This accessibility not only proved useful for learning the language itself, but also for aiding children in their first-language acquisition as well as encouraging them to learn other European languages, as the commonalities of these languages make understanding them easier⁵⁹ (Tellier 2013: 11).

Since its original formation, Esperanto has undergone many grammatical and lexical changes based on the conventions of its users, being codified only afterwards. By the time the second official dictionary was published in 1894, the number of lemmas had already doubled (Gobbo 2017). Words are mainly created based on the original agglutinative structure, although some words have been borrowed directly from other languages or changed into verbs by no general rule, and consequent vocabularies do not always coincide due to a lack of official regulation (Dyer 1923: 121–124). These irregularities, as well as some of Esperanto’s needlessly complex phonetic features like accented letters and diphthongs, led to the development of its descendent, Ido (an Esperanto suffix meaning ‘offspring’) in 1907 (De Beaufront 1919: xi). It is the product of seven years of work by the Delegation for the Adoption of an International Auxiliary Language, the members of which examined more than 60 schemes for an international language, and appointed a committee of representatives from various major linguistic groups to decide on an auxiliary language with improved regularity, facility, and internationality (Dyer 1923: 59). It retained most of Esperanto’s fundamental features with some improvements such as the simple grammatical forms (indicated with affixes), simple verb conjugation (sans exceptions), the practical agglutinative system (carried to a greater degree of logical precision), the practical use of compounding (similar to German), the grapheme-phoneme correspondence using only Roman numerals (without the use of diphthongs), and a more internationally accessible vocabulary (Dyer 1923: 78). Rather than being subject to the preferences of its speakers, the Ido vocabulary is controlled in that new additions need to be officially voted for by members of the Ido Academy (Dyer 1923: 121).

As a result, Ido is exceptionally regular, unambiguous, simplistic, and stable, yet is still aesthetically pleasing and comfortable to use. De Beaufront (1919: xi) holds it as having a “euphonious sound [...] something suggestive of Italian”. Despite all its benefits, however, Esperanto has gained a lot more public support. According to Ethnologue, it has an estimate of 2,001,000 speakers, of which roughly 1000 are first-language speakers (Simons and Fennig 2018). It is also available on Google Translate, and has its own Wikipedia edition. In contrast, Ido speakers are estimated around a mere few hundred (Blanke 2000), although no official census has been conducted for either language.

Although both Esperanto and Ido manage to mediate relatively well between the main European languages, in 1955 another artificial language was constructed in the hopes of being even more culturally neutral. Here we find Lojban, a continuation of the earlier Loglan (‘**logical language**’) project that attempted to construct a unique medium for interpersonal communication based on the principles of predicate logic. Its vocabulary was built using algorithms that mediated between the root words, sounds, and grammatical structures of six of the world’s most spoken languages, namely, Chinese, Hindi, English, Russian, Spanish, and Arabic (Nicholas and

⁵⁸ The syntactic roles of words are evident from their endings: nouns end in *-o*, adjectives end in *-a*, etc.

⁵⁹ For instance, the Esperanto sentence *La homo mangas panon* (‘The man eats the bread’) translates to *L’uomo mangia il pane* in Italian and *L’homme mange le pain* in French.

Cowan 2003: 6–7). Its primary purpose, however, was not to be a universal language, but to be a tool for studying and understanding language by removing as many constraints as possible that a language system may impose on clear thought and expression⁶⁰. It has a simple, regular, unambiguous morphology that allows for making fine distinctions between concepts with no exceptions to any rules, more so than that of Ido. The Lojban lexicon consists of 1350 roots that can easily be combined to form millions of words, and also uses an unambiguous phonetic spelling (Nicholas and Cowan 2003: 1–5).

Lojban is not only intended as a medium for interpersonal communication, but – owing to its formalised grammar which is similar to that of programming languages – also for potentially communicating with computers (Nicholas and Cowan 2003: 1–2). It consists of six vowels and 17 standard consonants of the Roman alphabet, and rarely makes use of capitalisation unless to indicate unusual stresses in names (Nicholas and Cowan 2003: 2). Otherwise, stresses are regular, falling on the penultimate syllable of a word unless the vowel is *y*, in which case it falls on the preceding syllable (Cowan 2016: 17). Its predicate grammar indicates the relationship between arguments (things, events, qualities, etc.) merely by their relative placement⁶¹. Consequently, the same arguments may serve a variety of syntactic roles depending on their order and the use of short structural words. The periods are used not to indicate the ends of sentences but rather serve as optional reminders for slight pauses between words so as to separate them phonetically (Nicholas and Cowan 2003: 2–4). Instead, punctuation is spoken as words: *i* is used to separate sentences, and *ni'o* is used to separate topics or paragraphs; multiple *ni'os* may also be used to separate sections (*ni'oni'o*) or chapters (*ni'oni'oni'o*) in longer texts (Cowan 2016: 21).

All of these features serve to ease the process of speech recognition and transcription (Nicholas and Cowan 2003: 2). As it contains no homonyms, polysemes or idioms, Lojban does not allow for wordplay, like puns, as ambiguous languages do. However, Nicholas and Cowan (2003: 6–8) suggest that it still allows for its own humorous spoonerisms, particularly through grammar manipulation, and that it is an exceptional medium for aphorisms. Moreover, its lucid structure and absence of cultural constraints makes it a particularly powerful tool for clear expression in abstract fields such as poetry, philosophy, physics, and metaphysics. Although its grammar is unambiguous, much of the disambiguating machinery in Lojban is optional: a speaker may still choose to omit portions of its logical structure to allow some ambiguity or vagueness (Nicholas and Cowan 2003: 6). The key difference is that, unlike in natural languages, being unspecific is a *choice* – “Your hearer may not understand what you meant: but will always understand what you said” (Nicholas and Cowan 2003: 5).

Each of these artificial languages successfully improves on some aspects of natural languages, although they also have their respective shortcomings. Some of those of Esperanto have already been mentioned, like the fact that its grammar and word-formations are not completely regular, that its vocabulary is not controlled, and that some of its vowels and diphthongs have complicated pronunciations – all of which are improved on by Ido. Another feature that Ido saw fit to change was the extensive (though not regular) use of *mal-* to indicate opposites as, although it reduces the number of root words, it requires “a sort of intellectual back somersault which is fatiguing and makes for clumsy diction that can be avoided by using the appropriate word” (Dyer 1923: 130). Both Esperanto and Ido can be critiqued by Lojban for being too

⁶⁰ This includes culture-specific constraints on one’s world view, drawn from the Sapir-Whorf hypothesis.

⁶¹ For instance, “Chris adores Pat” (*adores* being the relationship between the arguments *Chris* and *Pat*) can be translated as *la kris. prami la pat*, or even *la kris. la pat. prami* (Nicholas and Cowan 2003: 3).

Eurocentric. If the aim is to be culturally neutral, however, this also means Lojban may be a bit more difficult to learn for those familiar with common European natural language typologies and lexicons, and it will thus not have the added benefit of making those languages more easily intelligible. Considering language processing software, Lojban clearly wins as the best medium out of the three with its predicate grammar, although as yet it has the smallest number of active speakers⁶². Whether this is merely due to contingent factors or perhaps some inherent difficulty in using the language is unclear although, arguably, some of Lojban's disambiguating features (like the frequent pauses in sentences) might feel a bit unnatural/robotic in practice, at least at first. Overall, Ido seems to be the most balanced option of the three with regard to ease of acquisition and simplicity/regularity.

Regardless, what all of these examples serve to prove is that languages do not necessarily have to be as irregular, structurally complex, and ambiguous as those we are familiar with, that there is at least some interest in having an international auxiliary language, and that acquiring such a language can be a lot simpler than one might expect and is not necessarily at the cost of poetic and aesthetic merit. Such a language need not be limited to these three examples, but they do offer useful blueprints for future developments: perhaps finding a middle ground between the accessibility of Ido, the logicity of Lojban, with the regularity and simplicity of both. This is expressed well by Dyer (1923: 38):

The development of an I.L. [International Language] has been a matter of trial and error. No one man or group of men can sit down in a study and evolve a perfect form of language. It needs practical use to demonstrate its excellencies, its defects, its limits. The learning of the scholars must be checked by the common sense of the ordinary man.

This leads me to the final part of this subsection wherein I critically consider what may be gained by the adoption of such a constructed language in future as well as what may be lost.

5.2 Considerations for future attempts

5.2.1 What may be gained

The main objective of my discussion on artificial language is to consider how it may potentially aid human-robot interaction. Naturally, a formalised, simplistic, unambiguous language system like Ido or Lojban would overcome most of the difficulties of rule-based NLP that are touched on in the previous section, particularly those pertaining to parsing, disambiguation, and text generation. The value of a rule-based, rather than statistical, approach is that it could more easily allow for forms of unsupervised learning and address many of the issues associated with deep learning listed by Marcus (2018): it would be less data-heavy, solutions would be less superficial and should thus be more transferrable between different NLP applications, it would allow for greater transparency (and thereby control) of AI algorithms, it would be better at dealing with unfamiliar words if they are created and used according to specified rules (as in the case of Esperanto and Ido), and outcomes should be more predictable and reliable. Although it does not directly solve the problem of grounding the meaning of words in real-world phenomena⁶³, the

⁶² The number of active Lojban speakers is also unclear, but its online mailing list has no more than a couple hundred subscribers, of which less than a hundred are regular posters (Cowan 2016: 8).

⁶³ This might require a more embodied approach to AI language learning, i.e. mapping words to experiences.

formal grammar of an artificial language should make it significantly easier to understand the relationships between words (as in the unambiguous grammar of Lojban inspired by predicate logic). This, in turn, should make systems better at open-ended inferences, comprehensive text generation and summation, and common-sense reasoning.

Rather than spending all our time and resources on increasing the intricacy of our current statistical systems to deal with the complexities of natural language, the simplicity and clarity of an artificial language might allow us to avoid many of the difficulties completely, which means we could sooner focus on making our systems even more sophisticated from there⁶⁴. Moreover, if such a language ever becomes standardised as an international auxiliary language, the need for the particularly problematic field of MT may be significantly lessened. *If* we are ever able to develop some form of sapient⁶⁵ AI agents that are capable of expressing their experiences in language as we do, a formal, rational artificial language seems a more suited medium than one with a lot of arbitrary rules and exceptions. Furthermore, if we are ever to make use of AI computers as teachers, such a language may not only be easy to teach, but also to teach in.

There are of course many benefits to standardising such a formalised auxiliary language beyond the scope of computer applications. The most obvious, perhaps, is the benefits it would hold for its speakers, offering an easy medium of communication that allows for clearer expression and thereby less misunderstanding (Dyer 1923: 7). More importantly, by becoming an international medium, it may promote global unity by overcoming some communication barriers between nations as well as some antagonisms that may arise from cultural biases, misconception, and ignorance (Dyer 1923: 9, Nicholas and Cowan 2003: 9). This is especially useful since English, being the closest we currently have to an international language, takes years to master due to its particular complexities (De Beaufront 1919: vii). Although its grammar is relatively simple compared to that of other natural languages, its acquisition is slowed by its various irregularities and exceptions to rules, not to mention its chaotic spelling and pronunciation: the five vowels (*a, e, i, o, u*) correspond to five sounds in Spanish or Italian, while, in English, to more than 30 (Dyer 1923: 25). Not only does this give an unfair advantage to its native speakers, but, according to Nicholas and Cowan (2003: 7) and Li (2003: 36), as the universality of English is largely the result of centuries of imperial conquest and colonisation, its continued dominance serves to reinforce the cultural hegemonic dominance of the West.

For academic purposes, it may also prove useful as a neutral and easy-to-learn standard medium, “a central office where scholars and scientists of all nationalities can grab an idea fresh from the mint of thought” (Dyer 1923: 9–10). In addition, Nicholas and Cowan (2003: 9) and Dyer (1923: 13) maintain that the logical structure of an artificial language, free from the customs and biases of natural languages, forces speakers to be more conscious about their choice of words and thus formulate their arguments more carefully. From the moral side, Dyer (1923: 14) argues that this direct personal interchange between foreign speakers might also spark interest in other cultures

⁶⁴ This is not to say that symbolic systems should completely replace sub-symbolic approaches, as neural network processing remains an excellent tool for specific tasks like categorisation, recognition, prediction, and error correction, and the approaches may even be combined. It is also not the case that we will no longer require systems capable of processing natural languages, as the point of an auxiliary language is not to replace other languages, but is to be used as a medium alongside them.

⁶⁵ That is, having “a set of capacities associated with higher intelligence, such as self-awareness and being a reason-responsive agent” (Bostrom and Yudkowsky 2014: 323).

and may perhaps further language acquisition, and possibly foster a sense of kinship among humankind – all through a central medium that can be acquired in relatively no time at all.

5.2.2 What may be lost

A common concern regarding the standardisation of any particular language is that it may lead to a decline in the use of others, which would ultimately mean the loss of the unique cultural perspectives maintained by those languages (Waterlow 1913: 583). This also applies to the literary merit of various natural language texts that may become lost, even if translated. According to Dyer (1923: 88), this was one of the practical mistakes of the Esperantists, who had even attempted to translate Shakespeare – something he believes should only be attempted for texts like scientific or academic works that “have chief utility to the intellect, not to the sentiments” (Dyer 1923: 88). However, Li (2003: 52) emphasises that the whole point of an artificial language is to serve as an auxiliary, that is, a helpful addition for use in particular fields, and not a replacement for local languages.

Another common objection is that an artificial language would not be capable of the same aesthetic or literary merit as a natural one (Waterlow 1913: 583–584). As is made evident above, this is not necessarily the case – both Ido and Lojban are described as having an aesthetic phonetic quality akin to that of Italian due to their reliance on simple vowel sounds. Dyer (1923: 34) even argues that there is a different kind of beauty particular to a properly constructed language: “like the beauty that goes with the studied placing of the trees and the clearing away of the underbrush”. Moreover, as mentioned in the case of Lojban, poetry is possible in a formal language as it merely allows for a clearer expression of abstract concepts, and Nicholas and Cowan (2003: 8) contend that Lojban poets are already experimenting with new and existing forms of poetry that “seem especially well suited to the rhythm, sound, and flow of the language”.

Apart from such theoretical concerns, there are also various practical obstacles to implementing such a language. Li (2003: 42–43) suggests that the biggest of these is perhaps the fact that artificial languages, unlike natural ones, do not emerge spontaneously out of a real communicative necessity: “for a language to acquire and maintain its vitality, being able to use it for a broad range of meaningful communicative functions in natural settings is a prerequisite” (Li 2003: 43). This is also why he maintains that, after 125 years, Esperanto – despite all its benefits and large active community – has still not been able to exert enough influence to establish itself as a preferred international auxiliary language (Li 2003: 41). Another important factor in the spread of a language is political, economic and/or cultural pressure, which is largely determined by government support. Yet, Dyer (1923: 36–37) maintains that governments, as a rule, tend to be more conservative than the bulk of their people, and that it usually takes a long time to obtain official legislation after popular sentiment has built up. What may significantly speed up the process is the favourable verdict of the academic and scientific world, as it may build respectability and, in turn, lead to increased financial and social support (Dyer 1923: 38).

The idea of a constructed international auxiliary has been around for centuries, and seems to have much to offer as a medium in multiple fields. Although historic attempts have had limited successes, the possibility of natural and open communication between humans and AI technology/agents (that are actually able to effectively interpret what we say) might be the extra incentive needed to bring it into realisation.

6. Conclusion and final remarks

In this article, I consider some practical implications of the current trajectory of NLP technologies. We seem to be approaching a paradigm wherein all the more user interfaces and service operators take the form of AI software that we can communicate with, or at least give commands to, in natural language. Although some optimistic predictions suggest that we would be able to interact with these technologies in a way that feels organic (as if talking to another person), current NLP systems remain incapable of deep semantic analysis, effective and reliable translation, and generating complex text and rich and relevant answers to open-ended questions. It seems that a major obstacle is the fact that natural language, being the product of contingent and largely unregulated development, does not strictly adhere to formal rules, which means that its use and interpretation largely boils down to pragmatics, convention, and intuitive guessing. As a result, NLP and MT systems have had to rely mostly on surface-level statistical models rather than dealing with text on deeper, semantic levels.

Although recent developments in machine learning have significantly improved the quality of statistical systems, there may be inherent limitations to dealing with language merely on a surface level, and a growing consensus in the AI community suggests that deeper levels of symbolic abstraction and symbol-manipulation may be required for systems to reach human-level conversational ability. Rather than spending all our time and resources on making statistical systems better at dealing with the particular complexities of various natural languages, I question whether the process may not be more efficient if we give them a simpler medium to work with. For this, I consider two approaches: first, that we may simplify and regulate our current usage of English (having been the dominant focus of NLP research). Drawing from the natural simplification of English conventions by non-native speakers, as well as changes in communicative behaviour brought on by our increased use of digital communication platforms, I highlight some major areas in our language usage that are likely to be affected by increased interaction between us and AI systems: our choice of language, the complexity and length of our sentences, the size of our vocabulary, our use of borrowed terms, and our explicitness. Secondly, I consider the possibility of using a formalised artificial language as an auxiliary medium, as not only could it simplify language processing tasks and avoid many of the limitations of current statistical models, but it could also offer a more neutral, easy-to-learn additional language for uniting people from different linguistic backgrounds with none necessarily having the upper hand.

Drawing from existing examples, I find Esperanto and Ido, due to their origins in the major European languages, to have the dual benefit of making those languages slightly more comprehensible and easier to acquire for people who already speak those languages. On the other hand, as a more culturally-neutral language, Lojban has the benefit of enabling clear and logical expression of concepts without familiar cultural constraints, making it – at least in theory – an excellent medium for philosophy and poetry. Even if none of these proves feasible as they are, I postulate that they may offer helpful blueprints for developing a new artificial language that incorporates the best features of each.

Finally, I would like to offer a few closing remarks to clarify the scope of my argument. Firstly, as my argument is limited to current tendencies in NLP software, future developments in this area may lessen the need for simplifying and disambiguating language – ideally, it might even become possible to effectively process all major natural languages. Secondly, communicative

AI software may never become as ubiquitous as current predictions suggest, and if it does, it may never reach the point of achieving human-like, *natural* conversation. Thirdly, effective (human-like) NLP would ideally rely on a combination of other contextual factors such as verbal intonation, body language, eye movement, facial expression, etc., but for the purposes of this article, I focused specifically on language itself. However, finding ways to incorporate such factors may also significantly help disambiguation and semantic interpretation in future. Finally, this article does not attempt to make any normative claims about how language *should* be adapted to enable it to be processed successfully by computers, but is merely a consideration of possibilities that may allow us to overcome many of the inherent limitations of current NLP models, given our current trajectory of globalisation and communicative AI development.

References

- Ananiadou, S., J. McNaught and P. Thompson. 2012. *The English language in the digital age*. Heidelberg: Springer Nature.
- Bengio, Y. 2008. Neural net language models. *Scholarpedia* 3(1): 3881. <https://doi.org/10.4249/scholarpedia.3881>
- Benson, K. 2015. *Zero UI: The end of the screen-based interface*. Available online: <https://conferences.oreilly.com/solid/internet-of-things-2015/public/schedule/detail/40559> (Accessed 23 June 2018).
- Bentz, C., A. Verkerk, D. Kiela, F. Hill and P. Buttery. 2015. Adaptive communication: Languages with more non-native speakers tend to have fewer word forms. *PLOS One* 10(6): 1–23. <https://doi.org/10.1371/journal.pone.0128254>
- Bianzino, N.M. 2017. *Technology trends for people: AI is the new UI*. Available online: <https://www.wired.com/brandlab/2017/04/technology-trends-people-ai-new-ui/> (Accessed 23 June 2018).
- Blanke, D. 2000. Vom Entwurf zur Sprache. *Interface, Journal of Applied Linguistics* 15(1): 37–89.
- Bose, R. 2004. Natural language processing: Current state and future directions. *International Journal of the Computer, the Internet and Management* 12(1): 1–11.
- Bostrom, N. and E. Yudkowsky. 2014. The ethics of artificial intelligence. In K. Frankish and W. Ramsey (eds.) *The Cambridge handbook of artificial intelligence*. Cambridge: Cambridge University Press. pp. 316–334. <https://doi.org/10.1017/cbo9781139046855.020>
- Boyd, C. 2018. *The past, present, and future of speech recognition technology*. Available online: <https://medium.com/swlh/the-past-present-and-future-of-speech-recognition-technology-cf13c179aaf> (Accessed 5 September 2018).
- Bryson, B. 1990. *The mother tongue: English and how it got that way*. New York: William and Morrow Company.

- Clark, J.M. and E.A Gregor. 2012. *Global English (Globish) and its impact on the translator*. Available online: https://www.ata-chronicle.online/wp-content/uploads/4111_24_jeana_clark_esma_gregor.pdf (Accessed 18 August 2019).
- Cowan, J.W. 2016. *The complete Lojban language*. Available online: http://lojban.org/publications/cll/cll_v1.1_book.pdf (Accessed 25 July 2018).
- De Beaufront, L. 1919. *Complete manual of the auxiliary language Ido*. London: Sir Isaac Pitman and Sons.
- Dorr, B.J., E. Hovy and L. Levin. 2004. Machine translation: History of research and use. In K. Brown (ed.) *Encyclopedia of languages and linguistics*. Oxford: Elsevier. pp. 375–383.
- Dyer, L.H. 1923. *The problem of an international auxiliary language and its solution in Ido*. London: Sir Isaac Pitman and Sons.
- Gauthier, J. and I. Mordatch. 2016. A paradigm for situated and goal-driven language learning. *arXiv*: 1610.03585 [cs.CL]. Available online: <https://arxiv.org/pdf/1610.03585.pdf> (Accessed 18 August 2019).
- Gobbo, F. 2017. The lexicography of Esperanto. In P. Hanks and G.-M. de Schryver (eds.) *International handbook of modern lexis and lexicography*. Heidelberg: Springer. pp. 1–10. https://doi.org/10.1007/978-3-642-45369-4_88-1
- Goyal, P., S. Pandey and K. Jain, 2018. *Deep learning for natural language processing: Creating neural networks with Python*. New York: Apress.
- Gurney, K. 2004. *An introduction to neural networks*. London: UCL Press.
- Hartshorne, J.K. 2011. Where are the talking robots? *Scientific American Mind* 22(1): 44–51.
- Hickey, R. 2010. Language change. In M. Fried, J. Östman and J. Verschueren (eds.) *Variation and change: Pragmatic perspectives*. Amsterdam: John Benjamins. pp. 171–198.
- Hutchins, W.J. 2005. Example-based machine translation: A review and commentary. *Machine Translation* 19(3): 197–211. <https://doi.org/10.1007/s10590-006-9003-9>
- Hutchins, W.J. and H.L Somers. 1992. *An introduction to machine translation*. London: Academic Press.
- Karami, O. 2014. *The brief view on Google Translate machine*. Available online: <https://pdfs.semanticscholar.org/c6f9/5d543c0b34c4026b9e6cf64decd94b793823.pdf> (Accessed 18 August 2019).
- Kiela, D., L. Bulat, A.L. Vero and S. Clark. 2016. Virtual embodiment: A scalable long-term strategy for artificial intelligence research. *arXiv*: 1610.07432 [cs.AI]. Available online: <https://arxiv.org/pdf/1610.07432.pdf> (Accessed 18 August 2019).

- Koehn, P. 2017. Neural machine translation. *arXiv*: 1709.07809 [cs.CL]. Available online: <https://arxiv.org/pdf/1709.07809.pdf> (Accessed 18 August 2019).
- Li, D.C.S. 2003. Between English and Esperanto: What does it take to be a world language? *International Journal of the Sociology of Language* 164(2003): 33–63. <https://doi.org/10.1515/ijsl.2003.055>
- Luckhurst, P. 2018. *Say it: How millennials found their voice notes*. Available online: <https://www.standard.co.uk/tech/voice-notes-whatsapp-speech-millennials-a3796846.html> (Accessed 9 September 2018).
- Lucy, L. and J. Gauthier. 2017. Are distributional representations ready for the real world? Evaluating word vectors for grounded perceptual meaning. *arXiv*: 1705.11168 [cs.CL]. <https://doi.org/10.18653/v1/w17-2810>
- Madhavan, R. 2018. *Natural language processing: Current applications and future possibilities*. Available online: <https://www.techemergence.com/nlp-current-applications-and-future-possibilities/> (Accessed: 23 June 2018).
- Marcus, G. 2018. Deep learning: A critical appraisal. *arXiv*: 1801.00631 [cs.AI]. Available online: <https://arxiv.org/ftp/arxiv/papers/1801/1801.00631.pdf> (Accessed 18 August 2019).
- Mauranen, A. 2015. English as a global Lingua Franca: Changing language in changing global academia. In K. Murata (ed.) *Exploring ELF in Japanese academic and business contexts*. Amsterdam: John Benjamins. pp. 29–46.
- Nerrière, J.-P. 2003. Do you speak English or Globish? (Transl. N. Mellott). *Annales des Mines: Gérer et Comprendre* 2010/2(100): 56–64.
- Nicholas, N. and J.W. Cowan (eds.) 2003. *What is Lojban? .i la lojban. mo*. Available online: <https://www.bgcarlisle.com/blog/wp-content/uploads/2014/09/Lojban-Level-0.pdf> (Accessed 25 July 2018).
- Nield, T. 2019. *Is deep learning already hitting its limitations?* Available online: <https://towardsdatascience.com/is-deep-learning-already-hitting-its-limitations-c81826082ac3> (Accessed 25 June 2019).
- Omar, A. and M. Miah. 2012. Impact of technology on teens' written language. *International Journal of Advanced Trends in Computer Science and Engineering* 1(1): 9–17.
- Ramamurthy, R., A. Morya, L. Karthik, M. Vijay and A. Gupta. 2017. *The coming intelligent digital assistant era and its impact on online platforms*. Available online: <https://www.cognizant.com/whitepapers/the-coming-intelligent-digital-assistant-era-and-its-impact-on-online-platforms-codex2447.pdf> (Accessed 25 July 2018).
- Russell, S.J. and P. Norvig. 2010. *Artificial intelligence: A modern approach*. Upper Saddle River, NJ: Pearson Education.

- Siddharthan, A. 2006. Syntactic simplification and text cohesion. *Research on Language and Computation* 4(1): 77–109. <https://doi.org/10.1007/s11168-006-9011-1>
- Simons, G.F. and C.D. Fennig (eds.) 2018. *Ethnologue: Languages of the world*. Dallas: SIL International.
- Tellier, A. (ed.) 2013. *Esperanto as a starter language for child second-language learners in the primary school*. Berlaston: Esperanto UK. <https://doi.org/10.1075/lplp.37.1.12gre>
- Turing, A.M. 1950. Computing machinery and intelligence. *Mind* 49: 433–460.
- Waldrop, M.M. 1984. Natural language understanding. *Science* 224(4647): 372–374.
- Waterlow, S. 1913. “Interlingua” and the problem of universal language. *The Monist* 23(4): 567–585. <https://doi.org/10.5840/monist191323416>
- Watt, H.J. 2010. How does the use of modern communication technology influence language and literacy development? *Contemporary Issues in Communication Science and Disorders* 37: 141–148. https://doi.org/10.1044/cicsd_36_f_141
- Wit, E.J.C. and M. Gillette. 1999. *What is linguistic redundancy?* Chicago: University of Chicago.
- Wu, Y., M. Schuster, Z. Chen, Q.V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, Ł. Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. Corrado, M. Hughes and J. Dean. 2016. Google’s neural machine translation system: Bridging the gap between human and machine translation. *arXiv*: 1609.08144 [cs.CL]. https://doi.org/10.1162/tacl_a_00065
- Xuan, H.W., W. Li and G.Y. Tang. 2012. An advanced review of hybrid machine translation (HMT). *Procedia Engineering* 29(2012): 3017–3022. <https://doi.org/10.1016/j.proeng.2012.01.432>
- Young, T., D. Hazarik, S. Poria and E. Cambria. 2018. Recent trends in deep learning based natural language processing. *IEEE Computational Intelligence Magazine* 13(3): 55–75. <https://doi.org/10.1109/mci.2018.2840738>