



AI

MACHINE TRANSLATION AND
THE FUTURE OF LANGUAGE



A LOCARIA WHITEPAPER

Contents:

03 The Future of Translation

05 - So, how does the human brain produce language?

05 - How does AI translation differ?

06 - How do we distinguish the good translations from the bad?

07 - What does this mean for the future of AI translation?

08 Survey: Crossing Wires - Why AI and NLP are Important for Linguists

10 - Awareness of AI among professional linguists

10 - Impact on future work for professional linguists

11 - Where could automated translation solutions such SMT or NMT help the professional linguist in the future?

11 - How often do linguists currently use SMTs during their work?

12 - General thoughts from linguists about the future of their work

13 The Future of Linguistics

14 - Will humans keep learning languages?

15 - Will AI ever achieve human-level language production and be generally accepted for use?

AI has been receiving a significant amount of attention in the media lately as it finds more applications across a large number of industries.

This is especially exciting since improvements in neural machine translation (NMT) now mean that it's possible for numerous language combinations to get translations that are not only intelligible, but actually quite good. However, why is it still not perfect?

Why is it so difficult to generate natural, native-sounding language?

To understand this, we must consider how humans produce language as opposed to a machine.

THE FUTURE OF TRANSLATION

Humans create variations of words based on their interactions. Whether with friends, business contacts, family – we all do it. Unlike machines, we understand words from their overall context. Humans take numerous factors into account when forming a sentence: our knowledge about a subject, the words' relationship to our environment, and the company we're in at the time.

As we grow we continuously absorb images, developing emotions towards people and things that affect us. Consequently, we recognise sounds and voices and subconsciously associate them with those images, actions, and emotions.

‘We see something that triggers an emotional reaction, and subconsciously cross-reference it with our own language “database” to identify how best to express ourselves.’

So, how does the human brain produce language?

The older we get, the more talented we become at dissecting language. Our ability to express our thoughts coherently and efficiently improves. Experiences are recognised by our visual senses. These are then recorded and categorised. Emotions and intuition are triggered by chemicals that affect our nervous system. We may then use words to express those emotions.

We see something that triggers an emotional reaction, and subconsciously cross-reference it with our own language “database” to identify how best to express ourselves. These things must be experienced, and experience takes time to develop. There aren't any shortcuts. Simply remembering words isn't enough. It's what we associate with different variations of words that counts.

How does AI translation differ?

For machines such as NMT (neural machine translation) systems, the starting point is in the words and phrases that make up a source text, as well as any relevant content material in their database. This is what an NMT system's linguistic decisions for the best target text are based on, as opposed to the images, emotions and experiences that are a primary reference point for humans. While much can be achieved with highly advanced analysis of a source text, there are many

‘The expression “reading between the lines” refers to the drawing of inferences – not from words directly, but from our knowledge, or the context associated with a piece of text.’

environmental nuances which subconsciously guide the human linguistic decision-making process. How we express ourselves is impacted by implicit elements that we naturally incorporate into our language.

The expression “reading between the lines” refers to the drawing of inferences – not from words directly, but from our knowledge, experience, or the context associated with a piece of text. We can then use this to better express ideas or thoughts in another language when cultural nuances differ, or we are constrained by the structure of our mother tongue.

If future NMT solutions could consult enormous amounts of pre-existing, constantly updated, humanly translated data, and then analyse and incorporate surrounding variables from the speaking environment, only then could they produce human-level translations.

How do we distinguish the good translations from the bad?

One of the biggest challenges with machine translations is the on-going quality assessments they must undergo to ensure human standards are maintained. Numerous rating systems have been developed over recent years to judge the accuracy of SMTs (Statistical Machine Translations) and NMTs. The more common ones are: BLEU (Bilingual Evaluation Understudy), TER (Translation Edit Rate), and GTM (General Text Matcher).

Each has its pros and cons:

BLEU scoring splits text into segments, compares them to a corpus of existing human translations, and measures closeness using several statistical metrics.

TER supports linguists with post-editing. It uses the finalised target and its source text, compares the total text to existing, accepted translations in the MT, then provides the minimum number of edits required to optimise the target. This “edit rate” isn’t to be taken as an exact figure. Rather, it gives the linguist an idea of the required total effort. It may be that many of the edit suggestions aren’t serious or necessary, so this methodology isn’t ideal as a judge of the overall quality of an MT output.

GTM uses several similarity metrics which check for “hits” (two words that match in the candidate and reference text) and matches of “runs” (adjacent sequences of matching words). It does so across numerous variations, factoring in all identified matches regardless of length.

TER and GTM are said to function better than other rating systems, as they look at





numerous variations at different lengths, then provide a final metric indicating the effort required to improve the target text.

What does this mean for the future of AI translation?

For the foreseeable future, the difficulty any language scoring system has is functioning without a single objectively defined winning standard. There is no such thing as ‘the perfect, most accurate translation’. Anybody describing a translation positively in the superlative doesn’t understand that there are many accepted versions, all considered linguistically and possibly, stylistically correct.

‘There is no such thing as the perfect, most accurate translation’

For now, automatic MT evaluation measures are less reliable than human evaluations, and are still far from being able to entirely substitute human judgement. However, through the application of well-known evaluation metrics, MTs will continue to support linguists in getting a better idea of the quality of different translations — especially when used for standard text that must stick close to the source (e.g. technical manuals, instructions and descriptions).

Only by combining smart technology, as well as human input and decision-making, can a final version be produced within a minimal time frame at a reasonable price and, most importantly, at the expected level of quality.

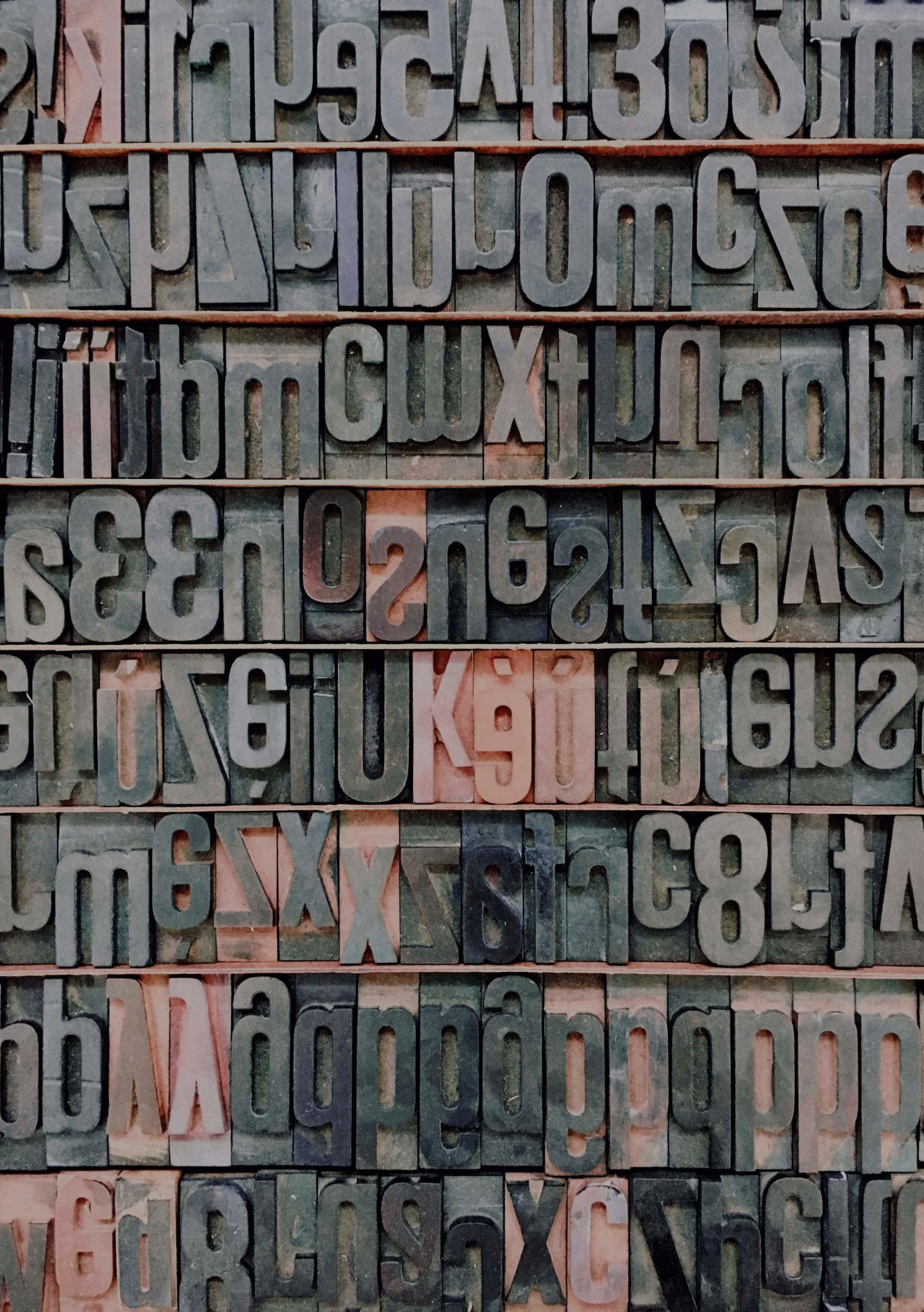
CROSSING WIRES

WHY AI AND NLP ARE IMPORTANT FOR LINGUISTS

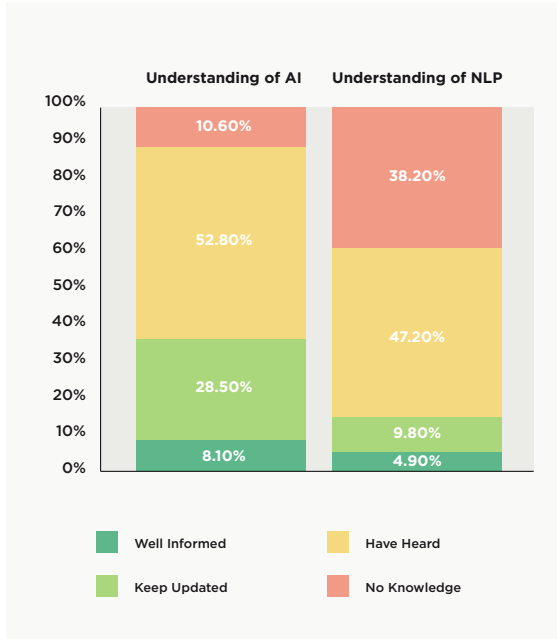
What do linguists make of AI and natural language processing (NLP)? Do they see a bright future for their careers with AI, or worry about being replaced by it entirely?

To find out, we ran a survey with 150 participating linguists from across the globe. The survey was a combination of questions that required them to select from a list of answers, or give their view in their own words.

An essential part of the survey saw each linguist describe their feelings towards AI, NLP, and machine translation (MT). This was fundamental in gauging the current mood. Let's take a closer look at the results.



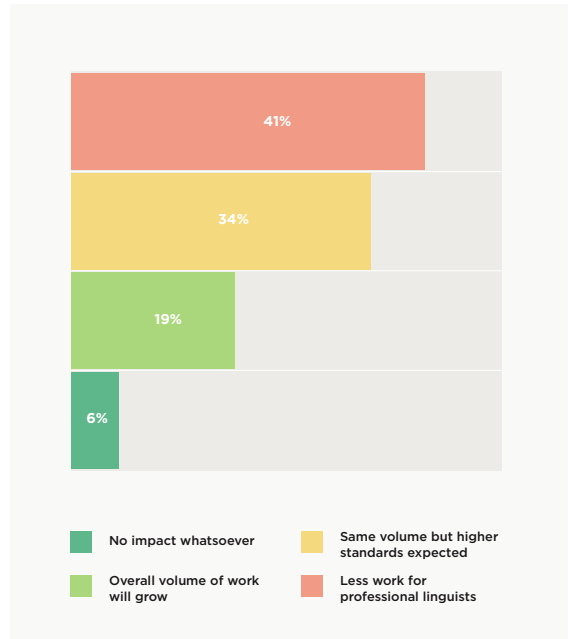
Awareness of AI among professional linguists



Considering the amount of publicity AI and MT solutions have received recently, the results here were surprising. Only about half of all participating linguists had heard about the technology.

Surprisingly, only about 28% of language professionals keep updated about AI. A shocking 38.2% have no knowledge whatsoever. While linguists are aware of MT tools — after all who hasn't heard of Google Translate? — there is an apparent lack of understanding of the connection between AI, NLP, and advanced MT such as statistical machine translation (SMT) and neural machine translation (NMT).

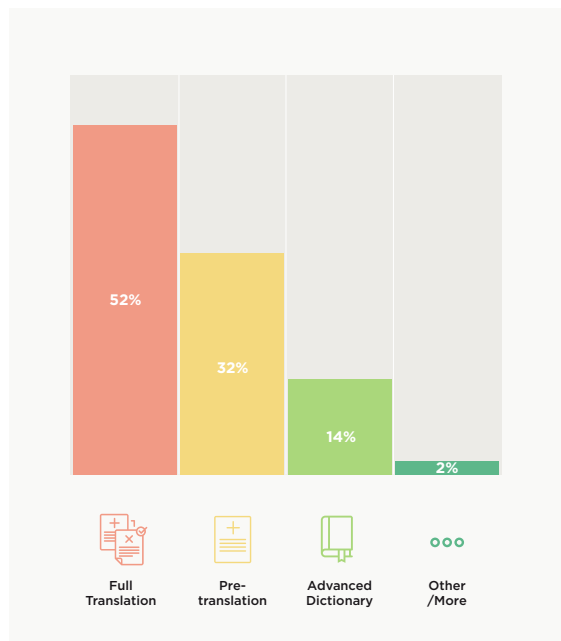
Impact on future work for professional linguists



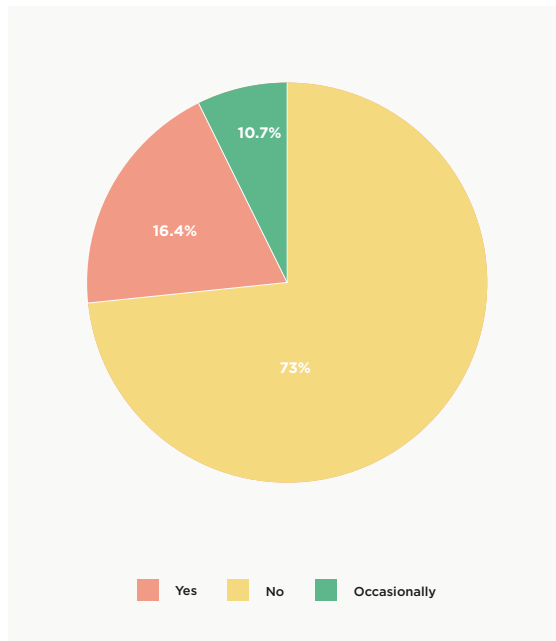
Almost half of the 150 linguists surveyed foresee a reduction in their work. Roughly a third believe they will receive the same amount of projects, but as their job becomes more focused on quality assurance (QA) and revision, final quality standards will significantly rise.

Interestingly, nearly a fifth of language professionals surveyed think that their work-load may grow with the increased usage of MT technology. This is likely, as content which isn't professionally translated for reasons of budget or time may be automatically pre-translated, while maintaining an accepted level of accuracy. As the demand for pre-translated MT content surges, the need for QA by professional linguists to add a human touch could also increase.

Where could automated translation solutions help the professional linguist in the future?



How often do linguists currently use SMTs during their work?

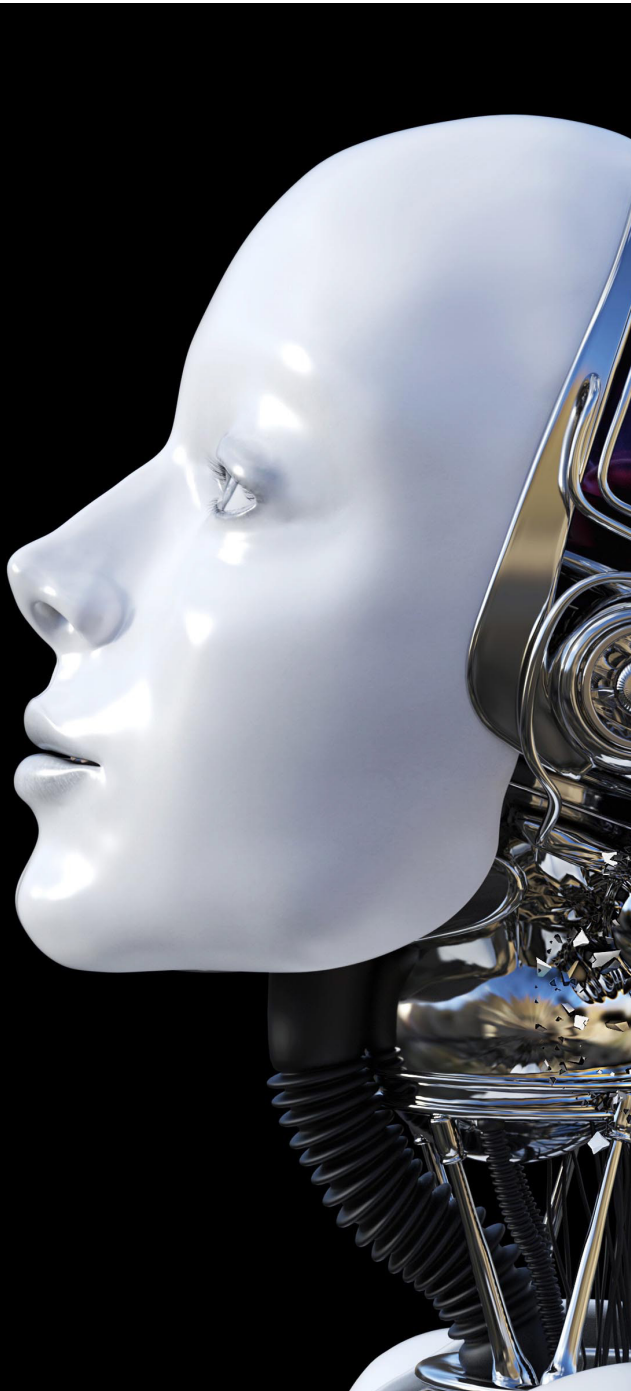


Once the usage and purpose of SMTs and NMTs are understood, most linguists start to think about how it could support them in their work instead of ignoring it as a threat to their livelihood.

SMT is already used by many to get a rough understanding of texts written in foreign languages. As confidence in the quality of SMTs grows, perhaps using them for professional purposes will gain more acceptance.

It's difficult to say how many linguists would admit to using SMTs for their translation projects, as despite improvements in SMT technology, its acceptance amongst clients is still low.

It's no surprise that 73% of the linguists surveyed have claimed that they never use SMTs during their work. A few years ago, that percentage was probably more like 99% since SMTs were still so new and unreliable. Therefore, the 16.4% of linguists who admit to using SMTs is expected to grow, especially in the wake of highly advanced NMT technology.



General thoughts from linguists about the future of their work

When asked to summarise their thoughts on AI and NLP, all seemed to agree on one key point — humans can never be fully replaced. There was an apparent split between two groups of people. There are those who see technology as an aid that could speed up the translation process. Meanwhile, many believe that people will begin to accept substandard language quality and adapt to a new reality in which content will be translated by machines.

‘When it comes to translation, humans can never be fully replaced’

We need to take a step back and understand how we produce and develop language. Where and how do our natural methodologies fundamentally differ from any machine-produced content?

At Locaria we very much doubt that the concerns of the latter group will materialise. Humans are excessively sensitive and protective of our languages. However, that doesn’t mean we won’t embrace AI translation. It instead depends on how intelligently we can weave technology into our natural language production processes to facilitate, or further enrich it.

THE FUTURE OF LINGUISTICS

Language isn't an exact science.

Often, there is no 'perfect' answer.

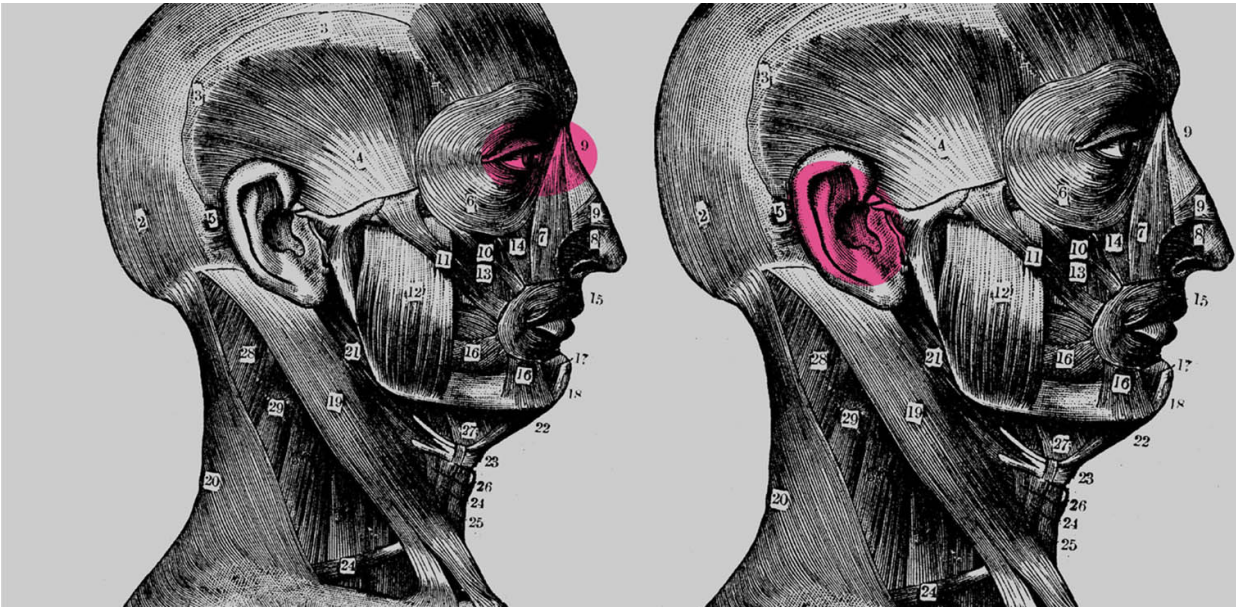
While much is logical, many elements are harder to explain, untethered as they are to any fixed set of rules. For instance,

when is a thought expressed with an indicative vs. a subjunctive mood? When to use polite vs. casual phrasing in languages such as Korean or Japanese?

How to articulate an expression that doesn't exist in a target language?

We often take language for granted.

Many people claim they speak their mother tongue perfectly, but who defines 'perfection'? AI has made us aware that technology can venture into anything, even challenge us in areas we consider intrinsically human.



AI compels us to enhance our language capabilities to maintain the edge and keep ahead of machine output. Machines challenge us in all areas, the realm of language is no exception. If future neural machine translation (NMT) solutions can continuously consult enormous amounts of constantly updated translated data, will language learning become obsolete? Perhaps with the aid of NMT and advanced AI, final translation will improve when not supported by any technology?

As technology tries to generate and translate natural language, we will become more sensitive towards what is right or wrong in a final piece of content. As machines begin reaching grammatical perfection, we may accept objective linguistic correctness, but be more critical of quality. We'll be more protective and purist about our languages. This is already the case in some instances, particularly in those languages and cultures heavily influenced by English and Western culture.

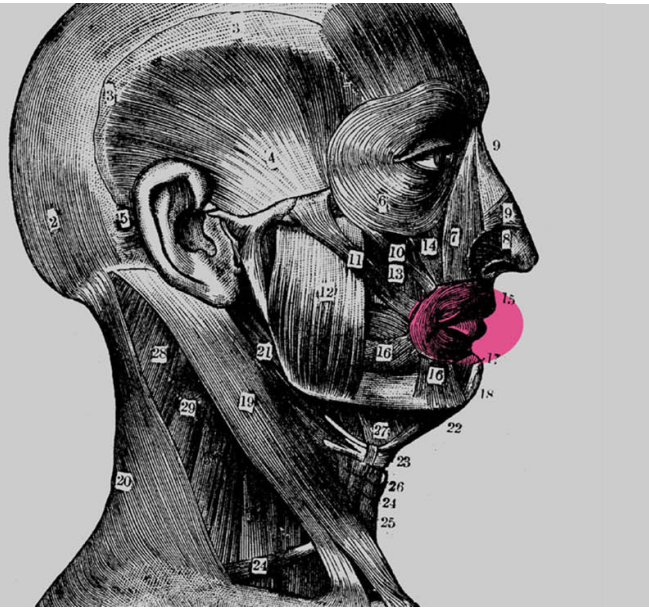
Will humans keep learning languages?

Technology has facilitated language learning. However, where vocabulary and grammar are

concerned, future AI may be hard to compete with. Its ability to quickly and efficiently absorb massive amounts of data and apply it effectively will always surpass human capability. However, humans will continue to have the edge in attaching context and emotions to words.

If AI can reliably translate any language content, who will bother learning the basics of a new language? Even through intensive study, a minimum of a year is usually needed to achieve some level of professional proficiency. Regardless of how hard you study, you must take breaks. Just like any muscle, the brain needs down-time to rest and process.

Allowing new and complex information to sink in and rewiring your brain takes time, regardless of your ability, and humans often pick the quickest, most convenient way of doing things. Sooner or later, the effort of acquiring a foreign language to professional standards may seem like an unsurmountable task. Especially if we must compete with what machines can provide at the press of a button. Further improvements in AI may see us focusing on our native languages — ensuring they



‘Humans will continue to have the edge in attaching context and emotions to words.’

evolve, and developing neologism to satisfy our ambition. However, this could mean that learning a second language will lose importance and become only a hobby for a few. Machine translation may achieve high standards which can only be topped by linguistic experts. Hence, only a few will pursue the mastery of a foreign language for leisure or business purposes.

Will AI ever achieve human-level language production and be generally accepted for use?

As we create, nurture, and develop our languages, AI may at some point create its own. These languages could originate from a combination of human languages it considers ideal to express thoughts and objectives. AI will have its own reasons for choosing one expression over another to best describe a situation. If no relevant existing word is found, it may create new words.

The tables may turn. We may study the definitions of words in AI languages, even incorporate them into our own. We may even translate between human and artificially created languages.

Regardless of how smart AI becomes, it will need systems and processes to pick up on-going changes in human languages and map them to wordings in its existing database. Future NMT solutions will only be able to match human-level translations once they learn to analyse context and draw from experience. This would involve a better understanding of the people involved in a conversation.

For the time being, machines still have a way to go. The processes of current NMT solutions are still far too narrow. To improve, AI must part with the flawed perception of language as but a sequential combination of words to create meaning. The winning technology will be one that can expand its analytical and learning capacity to understand the subconscious language decision-making process of humankind. We should be excited about the future of AI and machine translation.

A hybrid solution of NMT and humans will help protect and develop the richness of languages. Who knows, AI may even expand existing languages in ways we cannot yet imagine.

LOCARIA IS A PIONEERING DIGITAL LANGUAGE SERVICE PROVIDER, AND THE ONLY LSP CERTIFIED WITH THE WORLD'S MAJOR SEARCH ENGINES. WE COMBINE ENGAGEMENT METRICS WITH CONTENT CREATION PROCESSES BY HARNESSING DATA FROM ALL STAGES OF THE CUSTOMER JOURNEY INCLUDING AWARENESS, CONSIDERATION AND ENGAGEMENT. WE MAXIMISE THE EFFECTIVENESS OF GLOBAL CONTENT SEAMLESSLY ACROSS TERRITORY, PLATFORM OR MEDIA FOR BRANDS WITH INTERNATIONAL AMBITION. WE DO THIS IN MORE THAN 50 LANGUAGES, ACROSS MORE THAN 90 MARKETS. OUR GLOBAL CLIENTS CONSIST OF THE WORLD'S TOP BRANDS ACROSS LUXURY, RETAIL, TRAVEL, HOSPITALITY, FINANCE AND TECH.



Author:
Hannes Ben
Founder, Locaria

For further information,
please contact:
hello@locaria.com
www.locaria.com