THE DAWN OF THE HUMAN-MACHINE ERA

A FORECAST OF NEW AND EMERGING LANGUAGE TECHNOLOGIES

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

LANGUAGE IN THE HUMAN-MACHINE ERA

The Dawn of the Human-Machine Era

A forecast of new and emerging language technologies

LANGUAGE IN THE HUMAN-MACHINE ERA

This publication is based upon work from COST Action Language in the Human-Machine Era, supported by COST (European Cooperation in Science and Technology).

COST (European Cooperation in Science and Technology) is a funding agency for research and innovation networks. Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career and innovation.

www.cost.eu

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the Horizon 2020 Framework Programme of the European Union

To cite this report

Sayers, D., R. Sousa-Silva, S. Höhn et al. (2021). The Dawn of the Human-Machine Era: A forecast of new and emerging language technologies. Report for EU COST Action CA19102 'Language In The Human-Machine Era'. www.lithme.eu.

Contributors (names and ORCID numbers)

Sayers, Dave • 0000-0003-1124-7132
Sousa-Silva, Rui • 0000-0002-5249-0617
Höhn, Sviatlana • 0000-0003-0646-3738
Ahmedi, Lule • 0000-0003-0384-6952
Allkivi-Metsoja, Kais • 0000-0003-3975-5104
Anastasiou, Dimitra • 0000-0002-9037-0317
Beňuš, Štefan • 0000-0001-8266-393X
Bowker, Lynne • 0000-0002-0848-1035
Bytyçi, Eliot • 0000-0001-7273-9929
Catala, Alejandro • 0000-0002-3677-672X
Çepani, Anila • 0000-0002-8400-8987
Chacón-Beltrán, Rubén • 0000-0002-3055-0682
Dadi, Sami • 0000-0001-7221-9747
Dalipi, Fisnik • 0000-0001-7520-695X
Despotovic, Vladimir • 0000-0002-8950-4111
Doczekalska, Agnieszka • 0000-0002-3371-3803
Drude, Sebastian • 0000-0002-2970-7996
Fort, Karën • 0000-0002-0723-8850
Fuchs, Robert • 0000-0001-7694-062X
Galinski, Christian
Gobbo, Federico • 0000-0003-1748-4921
Gungor, Tunga • 0000-0001-9448-9422
Guo, Siwen • 0000-0002-6132-6093
Höckner, Klaus • 0000-0001-6390-4179
Láncos, Petra Lea • 0000-0002-1174-6882
Libal, Tomer • 0000-0003-3261-0180
Jantunen, Tommi • 0000-0001-9736-5425
Jones, Dewi • 0000-0003-1263-6332
Klimova, Blanka • 0000-0001-8000-9766
Korkmaz, Emin Erkan • 0000-0002-7842-7667
Maučec, Mirjam Sepesy • 0000-0003-0215-513X
Melo, Miguel • 0000-0003-4050-3473
Meunier, Fanny • 0000-0003-2186-2163
Migge, Bettina • 0000-0002-3305-7113
Mititelu, Verginica Barbu • 0000-0003-1945-2587
Névéol, Aurélie • 0000-0002-1846-9144
Rossi, Arianna • 0000-0002-4199-5898
Pareja-Lora, Antonio • 0000-0001-5804-4119
Sanchez-Stockhammer, C. • 0000-0002-6294-3579
Şahin, Aysel • 0000-0001-6277-6208
Soltan, Angela • 0000-0002-2130-7621
Soria, Claudia • 0000-0002-6548-9711
Shaikh, Sarang • 0000-0003-2099-4797
Turchi, Marco • 0000-0002-5899-4496
Yildirim Yayilgan, Sule • 0000-0002-1982-6609

A note on the contributors

This report began life in October 2020 at the start of the Language In The Human-Machine Era network (lithme.eu). Several online co-writing workshops followed, working together in Google Docs while video-conferencing. The list of contributors was recorded automatically in the Google Doc activity log. The content of the report was finalised on 12 May 2021, at which point this activity log was copied into a Google spreadsheet, and a 'table chart' automatically rendered to weigh contributions. On this basis LITHME's Chair, Dave Sayers, is the named first author. He is very closely followed in the activity log by Rui Sousa Silva, Chair of LITHME Working Group 1, and then by Sviatlana Höhn, LITHME's Vice-Chair. All three contributed significantly and consistently. The other named contributors all made the report what it is: authoritative, clear, diverse, and future-oriented. We look forward to working together on future editions of this important forecast.

1 Introduction: speaking through and to technology

"*Within the next 10 years, many millions of people will … walk around wearing relatively unobtrusive AR devices that offer an immersive and high-resolution view of a visually augmented world*" (Perlin 2016: 85)

The 'human-machine era' is coming soon: a time when technology is integrated with our senses, not confined to mobile devices. What will this mean for language?

Over the centuries there have been very few major and distinctive milestones in how we use language. The invention(s) of writing allowed our words to outlive the moment of their origin (Socrates was famously suspicious of writing for this reason). The printing press enabled faithful mass reproduction of the same text. The telegram and later the telephone allowed speedy written and then spoken communication worldwide. The internet enabled billions of us to publish mass messages in a way previously confined to mass media and governments. Smartphones brought all these prior inventions into the palms of our hands. The next major milestone is coming very soon.

For decades, there has been a growing awareness that technology plays some kind of active role in our communication. As Marshall McLuhan so powerfully put it, 'the medium is the message' (e.g. McLuhan & Fiore 1967; Carr 2020; Cavanaugh, Giapponi & Golden, 2016). But the coming human-machine era represents something much more fundamental. Highly advanced audio and visual filters powered by artificial intelligence – evolutionary leaps from the filters we know today – will overlay and augment the language we hear, see, and feel in the world around us, in real time, all the time. We will also hold complex conversations with highly intelligent machines that are able to respond in detail.

In this report we describe and forecast two imminent changes to human communication:

- **Speaking through technology**. Technology will actively contribute and participate in our communication – altering the voices we hear and facial movements we see, instantly and imperceptibly translating between languages, while clarifying and amplifying our own languages. This will not happen overnight, but it will happen. Technology will weave into the fabric of our language in real time, no longer as a supplementary resource but as an inextricable part of it.
- **Speaking to technology**. The current crop of smart assistants, embedded in phones, wearables, and home listening devices will evolve into highly intelligent and responsive utilities, able to address complex queries and engage in lengthy detailed conversation. Technology will increasingly understand both the content and the context of natural language, and interact with us in real time. It will understand and interpret what we say. We will have increasingly substantive and meaningful conversations with these devices. Combined with enhanced virtual reality featuring lifelike characters, this will increasingly enable learning and even socialising among a limitless selection of intelligent and responsive artificial partners.

In this introduction, we further elaborate these two features of the human-machine era, by describing the advance of key technologies and offering some illustrative scenarios. The rest of our report then goes into further detail about the current state of relevant technologies, and their likely future trajectories.

1.1 Speaking through technology

These days, if you're on holiday and you don't speak the local language, you can speak into your phone and a translation app will re-voice your words in an automated translation. This translation technology is still nascent, its reliability is limited, and it is confined to a relatively small and marketable range of languages.

The scope for error – and miscommunication, confusion or embarrassment – remains real. The devices are also clearly physically separate from us. We speak into the phone, awkwardly break our gaze, wait for the translation, and proceed in stops and starts. These barriers will soon fade, then disappear. In the foreseeable future we will look back at this as a quaint rudimentary baby step towards a much more immersive and fluid experience.

The hardware will move from our hands into our eyes and ears. Intelligent eyewear and earwear – currently in prototype – will beam augmented information and images directly into our eyes and ears. This is the defining distinction of the human-machine era. These new wearable devices will dissolve that boundary between technology and conversation. Our current binary understanding of humans on the one hand, and technology on the other, will drift and blur.

These devices will integrate seamlessly into our conversation, adding parallel information flows in real time. The world around us will be overlain by additional visual and audible information – directions on streets, opening hours on stores, the locations of friends in a crowd, social feeds, agendas, anything one could find using one's phone but instead beamed directly into one's eyes and ears. We will interact with machines imperceptibly, either through subtle finger movements detected by tiny sensors or through direct sensing of brainwaves (both are in development). This will alter the basic fabric of our interactions, fundamentally and permanently.

As these devices blossom into mass consumer adoption, this will begin to reshape the nature of face-to-face interaction. Instead of breaking the flow of conversation to consult handheld devices, our talk will be interwoven with technological input. We will not be speaking with technology, but *through* technology.

The software is also set to evolve dramatically. For example, the currently awkward translation scenario described above will improve, as future iterations of translation apps reduce error and ambiguity to almost imperceptible levels – finessed by artificial intelligence churning through vast and ever-growing databases of natural language. And this will be joined by new software that can not only speak a translation of someone's words, but automatically mimic their voice too.

Meanwhile, the evolution of Augmented Reality software, combined with emerging new eyepieces, will digitally augment our view of each person's face, in real time. This could alter facial movements, including lip movements, to match the automated voice translation. So we will hear people speaking our language, in their voice, and see their mouth move as if they were speaking those translated words. If our interlocutors have the same kit, they will hear

and see the same. This is what we mean when we say technology will become an active participant, inextricably woven into the interaction.

All this might feel like a sci-fi scenario, but it is all based on real technologies currently at prototype stage, under active development, and the subject of vast (and competing) corporate R&D investment. These devices are coming, and they will transform how we use and think about language.

1.2 Speaking to technology

As well as taking an active role in interaction between people, new smart technologies will also be able to hold complex and lengthy conversations with us. Technology will be the 'end agent' of communicative acts, rather than just a mediator between humans.

Currently, smart assistants are in millions of homes. Their owners call out commands to order groceries, adjust the temperature, play some music, and so on. Recent advances in chatbot technology and natural language interfaces have enabled people to speak to a range of machines, including stereos, cars, refrigerators, and heating systems.

Many companies use chatbots as a first response in customer service, to filter out the easily answerable queries before releasing the expense of a human operator; and even that human operator will be prompted by another algorithm to give pre-specified responses to queries. We already speak *to* technology, but in quite tightly defined and structured ways, where our queries are likely to fit into a few limited categories. This, too, is set to change.

New generations of chatbots, currently under active development, will not only perform services but also engage in significantly more complex and diverse conversations, including offering advice, thinking through problems, consoling, celebrating, debating, and myriad other topics. The change here will be in the volume and nature of conversation we hold with technology; and, along with it, the level of trust, engagement, and even emotional investment we develop.

Furthemore, devices will be able to solve complicated requests and find or suggest possible user intentions. This, too, will be entirely new terrain for language and communication in the human-machine era. Like the move to augmented reality eyewear and earwear, this will be qualitatively distinct from the earlier use(s) of technology.

Now switch from Augmented Reality to Virtual Reality, and imagine a virtual world of highly lifelike artificial characters all ready and willing to interact with us, on topics of our choice, and in a range of languages. Perhaps you want to brush up your Italian but you don't have the time or courage to arrange lessons or find a conversation partner. Would those barriers come down if you could enter a virtual world full of Italian speakers, who would happily repeat themselves as slowly as you need, and wait without a frown for you to piece together your own words? Language learning may be facing entirely new domains and learning environments.

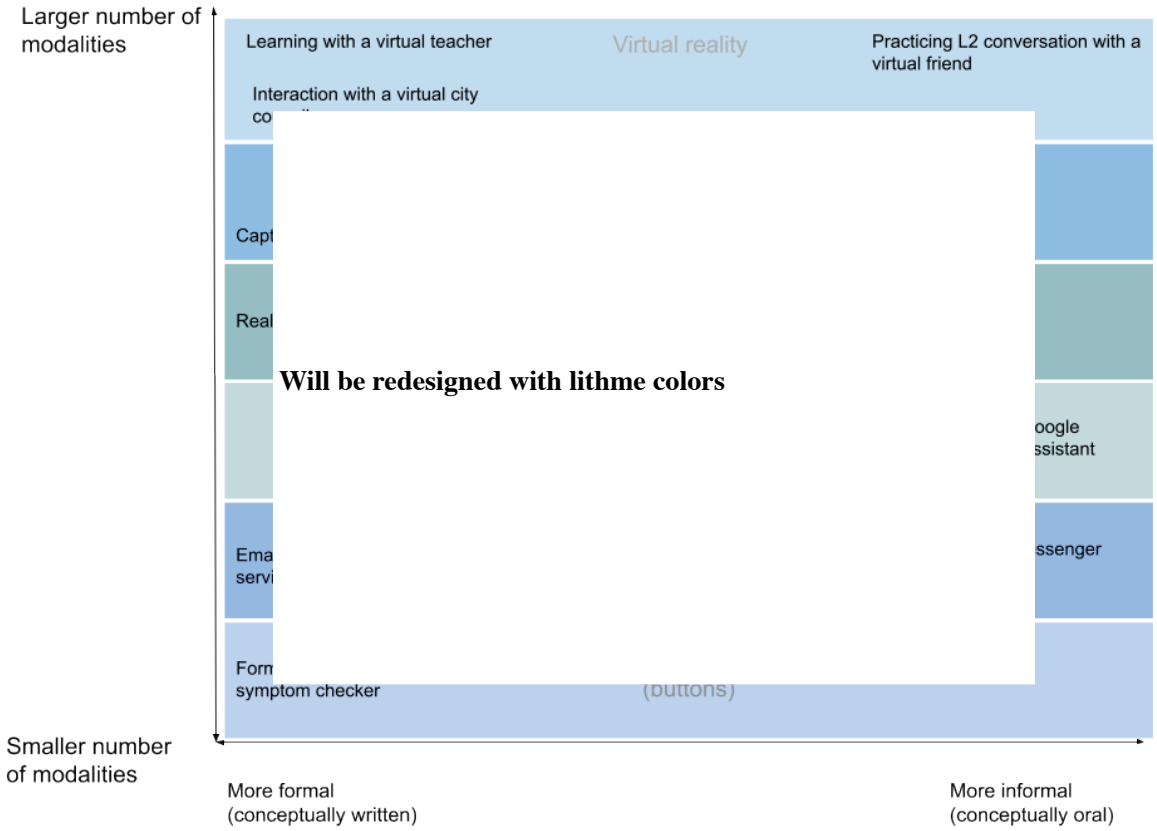
The same systems could be used for a range of other purposes, from talking therapy to coaching autistic children in interactional cues. The ability to construct a virtual world of lifelike interlocutors – who will never get scared or offended, never judge you, never laugh at you or gossip about you – carries with it immense potential for learning, training, and communication support. Indeed, highly intelligent chatbots are unlikely to remain constrained to specific contexts of use. They will adapt and learn from our input as silent algorithms contour their responses to maximise our satisfaction. As they become more widely available, many people may talk to them more or less all the time. Able to understand us, deploying algorithms to anticipate our needs, patiently responding and never getting tired or bored, bots may become our best imaginable friends.

Again, all this is simply a logical and indeed explicitly planned progression of current prototype technology, a foreseeable eventuality heading towards us. Many millions of people will soon be regularly and substantively speaking *to* technology.

1.3 The variety of languages, tools and use-cases

Below is a model that shows different levels of complexity in the different technologies we discuss in this report - from simple online form-filling to highly complex immersive Virtual Reality. We map two measures of complexity against each other: *formality*; and *number of modalities*. Formal language tends to be easier for machines to handle:

more predictably structured, with less variation and innovation. Informal language tends to be more free-flowing and innovative, harder to process. Next is modalities. Modalities are the various ways that humans use language through our senses, including writing, speech, sign, and touch. The more of these a machine uses at once, the more processing power is needed. The model below sets all these out for comparison.



There are predictions that over time the distinction between written and spoken language will gradually disappear, as more texts will be dictated to (and processed by) speech recognition tools, and texts that we read will become more speech-like.

Below we discuss types of human language, combining the perspectives of linguists and technologists. As above, this is relevant to the amount of work a machine must do.

1.3.1 Non-standard language (data)

Many languages around the world have a standard form (often associated with writing, education, and officialdom) alongside many non-standard varieties - dialects, and if the language is used internationally, perhaps also distinctive national varieties (for example Singaporean English or Morrocan Arabic). There will also be various registers of language, for example text messages, historical texts, formal letters, news media reporting, conversation, and so on (Biber & Conrad 2009). There will also be approximations associated with language learners.

All these variations present challenges for standard Natural Language Processing (NLP) methods, not least because NLP systems are typically trained on written, standard language such as newspaper articles. Usually, language processing with such language as input suffers from low accuracy and high rates of errors (Nerbonne 2016). Plank (2016) suggests "embracing" variations in linguistic data and combining them with proper algorithms in order to produce more robust language models and adaptive language technology.

Learner language is described as *non-standard* and *non-canonical language* in NLP literature because "learners tend to make errors when writing in a second language and in this regard, can be seen to violate the canonical rules of a language" (Cahill 2015). Other examples of non-canonical language are dialects, ordinary conversation and historical texts because they stray from the standard. Different approaches have been used to manage the contents of conversation with the user and to deal with learner errors. Wilske (2014) mentions constraining possible input and error diagnosis as strategies used by researchers and software developers in order to deal with the complexity of learner input.

1.3.2 Minority and under-resourced languages

Minority languages are typically spoken by a numerical minority in a given country or polity; languages such as Occitan or Sàmi. They tend to be under-resourced in terms of technology and the data needed for AI. Certain official languages of smaller countries face similar barriers, such as Latvian or Icelandic. Under-resourced languages suffer from a chronic lack of available resources (human-, financial-, time-, data- and technology-wise), and from the fragmentation of efforts in resource development. Their scarce resources are only usable for limited purposes, or are developed in isolation, without much connection with other resources and initiatives. The benefits of reusability, accessibility and data sustainability are often out of reach for such languages.

Until relatively recently, most research work in NLP has focused on just a few well-described languages each with abundant data. In fact, state-of-the-art NLP methodologies heavily rely on the availability of large amounts of data. However, the situation is rapidly evolving, as we discuss further in this report. Research and development are being driven both by a growing demand from communities, and by the scientific and technological challenges that this category of languages presents.

1.3.3 Sign languages

As discussed above, speech and writing are two modalities of language, two ways of transmitting meaning through human senses (hearing and sight respectively). There are other modalities, principally used by people with hearing and sight impairments, shown in Table 1.

'Sign languages' are those languages that typically use the signed modality. However, the above table risks some oversimplifications. Firstly, each 'sign language' is not simply a visual representation of e.g. English, Finnish, etc.; they are completely independent languages, with their own grammar, vocabulary, and other levels of linguistic structure. And, like spoken languages, they have huge amounts of variety, individual nuance, and creativity. Still, some spoken/written languages can be expressed using visual-spatial means, such as 'Signing Exact English' for expressing (spoken or written) English.

Modality	Meaning is encoded in...	Sense required	Commonly associated languages	Machine must produce...
Written	Graphemes (written characters)	Sight	English, Finnish, Esperanto, Quechua, etc.	Text
Spoken	Phonemes (distinctive sounds)	Hearing		Synthesised voice
Haptic	Touch (as in Braille or fingerspelling)	Touch		Moveable surface
Signed	Movements of the hands, arms, head and body; facial expression	Vision	British Sign Language, Finnish Sign Language, International Sign etc.	Avatar with distinguishable arms, fingers, facial features, mouth detail and posture

Table 1. *Modalities of language and what they require from machines*

Put another way, the signed modality is the basic modality for individual sign languages, but some other languages can also be expressed in the signed modality. It is possible to differentiate further into full sign languages and signed languages, such as fingerspelling, etc. often used in school education for young students (see ISO, in prep.). A further distinction is needed between visual sign languages and tactile sign languages. For example, unlike visual sign languages, tactile sign languages do not have clearly defined grammatical forms to mark questions. Additionally, visual sign languages use a whole range of visible movements beyond just the handshapes hearing people typically associated with sign. This includes facial expression, head tilt, eyebrow positions or other ways of managing what in spoken language would be intonation (Willoughby et al. 2018). "Unlike spoken languages, sign languages employ multiple asynchronous channels to convey information. These channels include both the manual (i.e. upper body motion, hand shape and trajectory) and non-manual (i.e. facial expressions, mouthings, body posture) features" (Stoll et al. 2018). It is important to distinguish all these, for understanding different people's needs and the different kinds of use cases of new and emerging language technologies.

1.3.4 Haptic language

The haptic modality is used particularly by deafblind people, who have limited or no access to the visual or auditory channels. Such communication systems can be based on an existing language (English, Finnish, etc.), often by adapting individual sign languages to the haptic modality or by fingerspelling in a spoken and written language. This may appear to be simply the use of the same language in a different modality; however, haptic systems are far more complicated. Deafblind signers have heterogeneous backgrounds and needs. For example, vision loss during life may lead to the development of idiosyncratic choices when language is developed in isolation. If a haptic system is not related to any other language but is instead an independent development, then it constitutes an individual language in its own right. Tadoma is a method of communication used by deafblind individuals, in which the deafblind person places their thumb on the speaker's lips and their fingers along the jawline. The middle three fingers often fall along the speakers cheeks with the little finger picking up the vibrations of the speaker's throat. See https://lifeprint.com/asl101/topics/tadoma.htm. (In the USA, the movements made by deafblind users to develop and promote interactional conventions have been referred to as 'pro-tactile movements' - see http://www.protactile.org/). 'Haptics', short for social-haptic communication, refers to a range of communicative symbols and practices that differ from standard tactile signing that are used to convey information, e.g. the description of a location, to deafblind people (Willoughby et al. 2018).

Braille is the written language used by blind people to read and write. It consists of raised dots corresponding to written characters, which can be 'read' with the fingers. Strictly speaking, communication through braille belongs to the haptic modality, although it is very close to writing, especially for the speaker. For extensive introductory detail on how Braille works, see e.g. http://www.dotlessbraille.org/.

For this report, a key detail is that there is not a one-to-one relationship between text in a visual alphabet and text in Braille. Even plain text needs to be translated into Braille before it can be read. To complicate matters further, Braille is language-specific, and the Braille code differs from country to country and according to domain (e.g., literary Braille, scientific Braille, Braille music, Braille poetry, pharmaceutical Braille), medium of rendition (six-dot Braille for paper, eight-dot for computers), and contraction levels (from two levels in British English Braille to five levels in the recently revitalised Norwegian Braille). Added to this comes the issue of Braille character sets (Christensen 2009).

In section 2.3, we discuss further the current capabilities and limitations of technologies for signed and haptic modalities.

1.4 Endless possibilities vs boundless risks, ethical challenges

The above scenarios sketch out some exciting advances, and important limitations. There are some additional conspicuous gaps in our story. Every new technology drags behind it the inequalities of the world, and usually contributes to them in ways nobody thought to foresee. Perhaps the most obvious inequality will be financial access to expensive new gadgets. This will inevitably follow - and perhaps worsen - familiar disadvantages, both enabling and disenfranchising different groups according to their means. Access will certainly not correlate to need, or environmental impact sustained (Bender et al. 2021).

There have already been concerns raised about inequalities and injustice in emerging language technologies, for example poorer performance in non-standard language varieties (including of ethnic minorities), or citizens being unjustly treated due to technologies (https://www.dailydot.com/debug/facebook-translation-arrest/). NLP is widely used to support decisions in life-altering scenarios including employment, healthcare (Char et al. 2018), justice, and finance: who gets a loan, who gets a job, who is potentially a spy or a terrorist, who is at risk of suicide, which medical treatment one receives, how long a prison sentence one serves, etc. But NLP is trained on human language, and human language contains human biases (Saleiro et al. 2020). This inevitably feeds through into NLP tools and language models (Blodgett, Barocas, Daumé & Wallach 2020). Work is underway to address this (Bender 2019; Beukeboom & Burgers 2020; Benjamin 2020; Saleiro et al. 2020). Remedies could lead to improved equality, or perhaps polarise society in new ways. LITHME is here to pay attention to all these possible outcomes, and to urge collaboration that is inclusive and representative of society.

A further major gap was discussed in the previous section: sign languages. There have been many attempts to apply similar technology to sign language: 'smart gloves' that decode gestures into words and sentences, and virtual avatars that do the same in reverse. But the consensus among the deaf community so far is that these are a profoundly poor substitute for human interpreters. They over-simplify, they elide crucial nuance, and they completely miss the diversity of facial expression, body posture, and social context that add multiple layers of meaning, emphasis and feeling to sign. Moreover, these technologies help non-signers to understand something from signs but they strip signers of much intended meaning. The inequality is quite palpable. There are early signs of progress, with small and gradual steps towards multimodal chatbots which are more able to detect and produce facial movements and complex gestures. But this is a much more emergent field than verbal translation, so for the foreseeable future, sign language automation is distantly inferior.

Another issue is privacy and security. The more we speak through and to a company's technology, the more data we provide. AI feeds on data, and improves by learning from our behaviour, from our data. We already trade privacy for technology. AI, the Internet of Things and social robots all offer endless possibilities, but they may conceal boundless risks. Whilst improving user experiences, reducing health and safety risks, easing communication between languages and other benefits, technology can also lead to discrimination and exclusion, surveillance, and security risks. This can take many forms. Some exist already, and may be exacerbated, like the "filter bubbles" (Pariser 2011), "ideological frames" (Scheufele, 1999; Guenther, Ruhrmann et al. 2020) or "echo chambers" (Cinelli et al., 2021) of social media, which risk intellectual isolation and constrained choices (Holone 2016). Meanwhile automatic text generation will increasingly help in identifying criminals based on their writing, for example grooming messages or threatening letters, or a false suicide letter. Such text generation technologies can also challenge current plagiarism detection methods and procedures, and allow speakers and writers of a language to plagiarise other original texts. Likewise, the automatic emulation of someone's speech can be used to trick speech recognition systems used by banks, thus contributing to cybercriminal activities. New vectors for deception and fraud will emerge with every new advance.

The limits of technology must be clearly understood by human users. Consider the scenario we outlined earlier, or a virtual world of lifelike characters - endlessly patient interlocutors, teachers, trainers, sports partners, and plenty else besides. Those characters will never be truly sad or happy for us, or empathise - even if they can emulate these things. We may be diverted away from communicating and interacting with - imperfect but real - humans.

Last but not least, another challenging setting for technology is its use by minority languages communities. From a machine learning perspective, the shortage of digital infrastructure to support these languages may hamper development of appropriate technologies. Speakers of less widely-used languages may lag in access to the exciting resources that are coming. The consequences of this can be far-reaching, well beyond the technological domain: unavailability of a certain technology may lead speakers of a language to use another one, hastening the disappearance of their language altogether.

LITHME is here to scrutinise these various critical issues, not simply shrug our shoulders as we cheer exciting shiny new gadgets. A major purpose of this report, and of the LITHME network, is to think through and foresee future societal risks as technology advances, and amplify these warnings so that technology developers and regulators can act pre-emptively.

1.5 The way ahead

LITHME is a diverse network of researchers, developers and other specialists, aiming to share insights about how new and emerging technologies will impact interaction and language use. We hope to foresee strengths, weaknesses, opportunities and threats. The remainder of this report sketches the likely way ahead for the transformative technologies identified above.

We move on now to a more detailed breakdown of new and emerging language technologies likely to see widespread adoption in the foreseeable future. The rest of the report falls into two broad areas: software; and hardware. Section 2 examines developments in computing behind the scenes: advances in Artificial Intelligence, Natural Language Processing, and other fields of coding that will (em)power the human-machine era. Section 3 focuses on the application of this software in new physical devices, which will integrate with our bodies and define the human-machine era.

2 Behind the scenes: the software powering the human-machine era

Summary and overview

Artificial Intelligence (AI) is a broad term applied to computing approaches that enable machines to 'learn' from data, and generate new outputs that were not explicitly programmed into them. AI has been trained on a wide range of inputs, including maps, weather data, planetary movements, and human language. The major overarching goal for language AI is for machines to both interpret and then produce language with human levels of accuracy, fluency, and speed.

Recent advances in 'Neural Networks' and 'deep learning' have enabled machines to reach unprecedented levels of accuracy in interpretation and production. Machines can receive text or audio inputs and summarise these or translate them into other languages, with reasonable (and increasing) levels of comprehensibility. They are not yet generally at a human level, and there is distinct inequality between languages, especially smaller languages with less data to train the AI, and sign languages - sign is a different 'modality' of language in which data collection and machine training are significantly more difficult.

There are also persistent issues of bias. Machines learn from large bodies of human language data, which naturally contains all of our biases and prejudices. Work is underway to address this ongoing challenge and attempt to mitigate those biases.

Machines are being trained to produce human language and communicate with us in increasingly sophisticated ways - enabling us to talk *to* technology. Currently these chatbots power many consumer devices including 'smart assistants' embedded in mobile phones and standalone units. Development in this area will soon enable more complex conversations on a wider range of topics, though again marked by inequality, at least in the early stages, between languages and modalities.

Automatic recognition of our voices, and then production of synthesised voices, is progressing rapidly. Currently machines can receive and automatically transcribe ~~language in~~ many languages, though only after training on several thousand hours of transcribed audio data. This presents issues for smaller languages.

Deep learning has also enabled machines to produce highly lifelike synthetic voices. Recently this has come to include the ability to mimic real people's voices, based on a similar principle of churning through long recordings of their voice and learning how individual sounds are produced and combined. This has remarkable promise, especially when combined with automated translation, for both dubbing of recorded video and translation of conversation, potentially enabling us to talk in other languages, in our own voice. There are various new ways of talking *through* technology that will appear in the coming years.

Aside from text and voice, attempts are underway to train AI on sign language. Sign is an entirely different system of language with its own grammar, and uses a mix of modalities to achieve full meaning: not just shapes made with the hands but also facial expression, gaze, body posture, and other aspects of social context. Currently AI is only being trained on handshapes; other modalities are simply beyond current technologies. Progress on handshape detection and production is focused on speed, accuracy, and making technologies less intrusive - moving from awkward sensor gloves towards camera-based facilities embedded in phones and webcams. Still, progress is notably slower than for the spoken and written modalities~~W~~.

A further significant challenge for machines will be to understand what lies beyond just words, all the other things we achieve in conversation: from the use of intonation (questioning, happy, aggressive, polite, etc.), to the understanding of physical space, implicit references to common knowledge, and other aspects woven into our conversation which we typically understand alongside our words, almost without thinking, but which machines currently cannot.

Progress to date in all these areas has been significant, and more has been achieved in recent years than in the preceding decades. However, significant challenges lie ahead, both in the state of the art and in the equality of its application across languages and modalities.

This section covers advances in software that will power the human-machine era. We describe the way machines will be able to understand language. We begin with text, then move on to speech, before looking at paralinguistic features like emotion, sentiment, and politeness.

Underlying these software advances are some techniques and processes that enable machines to understand human speech, text, and to a lesser extent facial expression, sign and gesture. 'Deep learning' techniques have now been used extensively to analyse and understand text sequences, to recognise human speech and transcribe it to text, and to translate between languages. This has typically relied on 'supervised' machine learning approaches; that is, large manually annotated corpora from which the machine can learn. An example would be a large transcribed audio database, from which the machine could build up an understanding of the likelihood that a certain combination of sounds correspond to certain words, or (in a bilingual corpus) that a certain word in one language will correspond to another word in another language. The machine learns from a huge amount of data, and is then able to make educated guesses based on probabilities in that data set.

The term 'Neural Networks' is something of an analogy, based on the idea that these probabilistic models are working less like a traditional machine - with fixed inputs and outputs - and more like a human brain, able to arrive at new solutions somewhat more independently, having 'learned' from prior data. This is a problematic and somewhat superficial metaphor; the brain cannot be reduced to the sum of its parts, to its computational

abilities (see e.g. Epstein, 2016; Cobb, 2020; Marincat, 2020). Neural Networks do represent a clear advance from computers that simply repeated code programmed into them. Still, they continue to require extensive prior data and programming, and have less flexibility in computing the importance and accuracy of data points. This is significant in the real world because, for example, the large amounts of data required for deep learning are costly and time consuming to gather. Investment has therefore followed the line of greatest utility and profit with lowest initial cost. Low-resource languages lose out from deep learning.

'Deep Neural Networks' (DNNs), by contrast, work by building up layers of knowledge about different aspects of a given type of data, and establishing accuracies more dynamically. DNNs enable much greater flexibility in determining, layer by layer, whether a sound being made was a 'k' or a 'g' and so on, and whether a group of sounds together corresponded to a given word, and words to sentences. DNNs allow adaptive, dynamic, estimated guesses of linguistic inputs which have much greater speed and accuracy. Consequently, many commercial products integrate speech recognition; and some approach a level comparable with human recognition.

Major recent advances in machine learning have centred around different approaches to ~~neural networks~~. Widely used technical terms include Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRUs). Each of these three can be used for a technique known as sequence-to-sequence, 'seq2seq'. Introduced by Google in 2014 (https://arxiv.org/pdf/1409.3215.pdf), seq2seq analyses language input (speech, audio etc.) not as individual words or sounds, but as combined sequences; for example in a translation task, interpreting a whole sentence in the input (based on prior understanding of grammar) and assembling that into a likely whole sentence in a target language - all based on probabilities of word combinations in each language. This marks a major advance from translating word for word, and enables more fluent translations. In particular it allows input and output sequences of different lengths, for example a different number of words in the source and translation - useful if source and target languages construct grammar differently (for example presence of absence of articles, prepositions, etc.) or have words that don't translate into a single word in another language.

The above is a highly compressed review of some of the underlying machinery for machine learning of language. Worth also noting that many of these same processes are used in areas like automatic captioning of photos (interpreting what is in a photo by comparing similar combinations of colours and shapes in billions of other photos), facial recognition (identifying someone's unique features by referring to different 'layers' of what makes a face look like a human, like a man, like a 45 year old, and so on), self-driving cars (distinguishing a cyclist from a parking space), and so on. These algorithms will govern far more than language technology in the human-machine era.

We move on now to discuss how these underlying machine smarts are used to analyse text, speech, paralinguistic features like sentiment, and then visual elements like gesture and sign.

2.1 Text Technology

Headline terminology for automated text facilities include: information extraction, semantic analysis, sentiment analysis, machine translation, text summarisation, text categorisation, keyword identification, named entity recognition, and grammar/spell-checkers, among others. A major challenge for NLP research is that most information is expressed as unstructured text. Computational models are based on numerical entities and probabilistic modelling; but natural language is obviously not so straightforward. Furthermore, the number of categories that exist in natural language data is magnitudes greater than, say, image processing. Success in NLP applications has therefore been slower and more limited.

2.1.1 Translation of texts

Humans have long had high hopes for machine translation; but for many years these hopes were in vain. The ALPAC report (Pierce & Carroll 1966) conveyed a ~~significant~~ sense of that disappointment. Significant technological investment at this time was paying off in the developments of the early internet. Investment in machine translation, however, generated much less satisfying results.

Initial attempts at machine translation were rule-based, built on the assumption that, if a computer was given a set of rules, eventually it would be able to translate any combination of words. Preliminary results of trials run on short messages produced under tightly controlled circumstances were promising. However, when fed texts pro-

duced naturally (often containing ungrammatical formulations), the system fell down. This is because translation is not about words, but about meanings. Computers have long struggled to process meanings in a source language and ~~process~~ them in a target language.

Attempts at machine translation were soon dropped, but were resumed later on by projects such as Google Translate, which approached the problem not based on rules but statistics, not on direct dictionary correspondence but on the likelihood of one word following another, or surrounding others in the semantic space. Statistical machine translation systems first aligned large volumes of text in a source and target language side by side, and then arrived at statistical assumptions for which words or word combinations were more likely to produce the same meanings in another language. Companies like Google were ideally placed for this, as they indexed trillions of pages written in many languages. The system would soon become a victim of its own success, as companies and users worldwide started using poor quality translations, including those produced by Google, to produce websites in many different languages. As a result, poor quality data fed into the same system. Garbage in, garbage out. Statistical machine translation, too, then fell short of expectations, and Google invited their users to correct the translations produced by the system.

Translation is nowadays perhaps the area where human-machine interaction technologies have advanced the most. Yet, not all types of translation have evolved at the same pace; translation of written language has progressed more than spoken and haptic languages.

More recently, research has focused on neural machine translation (NMT). The rationale behind NMT is that technology is able to simulate human reasoning and hence produce human-like machine translations. Indeed, the functions of MT are likely to continue to expand. In the area of machine translation there are now various utilities including Google Translate (https://translate.google.com/), Microsoft Translate (https://www.bing.com/translator) and DeepL (https://www.deepl.com/translator). Open source alternatives include ESPNet, and FBK-Fairseq-ST.

These are based on deep learning techniques, and can produce convincing results for many language pairs. Deep learning uses large datasets of previously translated text to build probabilistic models for translating new text. There are many such sources of data. One example is multilingual subtitles: and within these, a particularly useful dataset comes from TED talks - these are routinely translated by volunteers into many languages with administratively managed quality checks; they cover a variety of topics and knowledge domains, and they are open access (Cettolo et al. 2012). There are limitations, for example translations are mainly from English to other languages; and since many talks are pre-scripted, they may not represent typical conversational register (Dupont & Zufferey 2017; Lefer & Grabar 2015). TED talks are nevertheless valuable for parallel data. They are employed as a data set for statistical machine translation systems and are one of the most popular data resources for multilingual neural machine translation (Aharoni et al. 2019; Chu et al. 2017; Hoang et al. 2018; Khayrallah et al. 2018; Zhang et al. 2019).

The accuracy of machine translation is lower in highly inflected languages (as in the Slavic family), and agglutinative languages (like Hungarian, Turkish, Korean, and Swahili). In many cases, this can be remedied with more data, since the basis of deep learning is precisely to churn through huge data sets to infer patterns. This, however, presents problems for languages spoken by relatively small populations - often minority languages. Hence, progress is running at different paces, with potential for inequalities.

Even though deep learning techniques can provide good results, there are still rule-based machine translation systems in the market like that of the oldest machine learning company SYSTRAN (www.systransoft.com). There are also open source machine translation systems like Apertium (www.apertium.org). These toolkits allow users to train neural machine translation (NMT) systems with parallel corpora, word embeddings (for source and target languages), and dictionaries. The different toolkits offer different (maybe overlapping) model implementations and architectures. Nematus (https://github.com/EdinburghNLP/nematus) implements an attention-based encoder-decoder model for NMT built in Tensorflow. OpenNMT (https://opennmt.net/, https://www.aclweb.org/anthology/P17-4012) and MarianNMT (https://marian-nmt.github.io/) are two other open source translation systems. One of the most prolific open source machine translation systems is the Moses phrase-based system (www.statmt.org/moses), used by Amazon and Facebook, among other corporations. Moses was also successfully used for translation of MOOCs across four translation directions – from English into German, Greek, Portuguese, and Russian (Castilho et. al. 2017).

Another research trend is AI-powered Quality Estimation (QE) of machine translation. This provides a quality

indication for machine translation output without human intervention. Much work is being undertaken on QE, and some systems such as those of Memsource (https://www.memsource.com/features/translation-quality-estimation/) are available; but so far none seems to have reached sufficient robustness for large-scale adoption. According to Sun et al. (2020), it is likely that QE models trained on publicly available datasets are simply *guessing* translation quality rather than estimating it. Although QE models might capture fluency of translated sentences and complexity of source sentences, they cannot model adequacy of translations effectively. There could be various reasons for this, but this ineffectiveness has been attributed to potential inherent flaws in current QE datasets, which cause the resulting models to ignore semantic relationships between translated segments and the originals, resulting in incorrect judgments of adequacy.

CJEU MT Systran – SYStem TRANSlation has contributed significantly to machine translation (https://curia.europa.eu/jcms/upload/docs/application/pdf/2013-04/cp130048en.pdf). Another example is the European Union's eTranslation online machine translation service, which is provided by the European Commission (EC) for European official administration, small and medium sized enterprises (SMEs), and higher education institutions (https://ec.europa.eu/info/resources-partners/machine-translation-public-administrations-etranslation_en). ~~The~~ Bergamot ~~project~~ (https://browser.mt/) is a further interesting project whose aim is to add and improve client-side machine translation in a web browser. The project will release an open-source software package to run inside Mozilla Firefox. It aims to enable bottom-up adoption by non-experts, resulting in cost savings for private and public sector users. Lastly, ParaCrawl (https://paracrawl.eu/) is a European project which applies state-of-the-art neural methods to the detection of parallel sentences, and the processing of the extracted corpora.

As mentioned above, translation systems tend to focus on languages spoken by large populations. However, there are systems focusing on low-resource languages. For instance, the GoURMET project (https://gourmet-project.eu/) aims to use and improve neural machine translation for low-resource language pairs and domains. The WALS database (https://wals.info/) (Dryer and Haspelmath 2013) is used to improve systems (language transfer), especially for less-resourced languages (Naseem et al. 2012; Ahmad et al. 2019).

Machine translation has been particularly successful when applied to specialized domains, such as education, health, and science. Activities focused on specific domains abound: for example, the Workshop for Machine Translation (WMT) has offered a track on biomedical machine translation which has led to the development of domain-specific resources. http://www.statmt.org/wmt20/biomedical-translation-task.html. There are limited parallel corpora, and much more monolingual data in specialized domains (e.g., for the biomedical domain: https://www.aclweb.org/anthology/L18-1043.pdf). Back-translation is studied to integrate monolingual corpus into NMT training of domain-adapted machine translation (https://www.aclweb.org/anthology/P17-2061.pdf).

European Language Resource Coordination (ELRC) — http://lr-coordination.eu/node/2 — is gathering data (corpora) specialised on Digital Service Infrastructures. The EU's Connecting Europe Facility (CEF) in Telecom enables cross-border interaction between organisations (public and private). Projects financed by CEF Telecom usually deliver domain-specific corpora (especially for less resourced languages) for training and tuning of the e-Translation system. Examples of such projects include MARCELL and CURLICAT.

Currently, the main obstacle is the need for huge amounts of data. As noted above, this creates inequalities for smaller languages. Current technology based on neural systems conceal a hidden threat: neural systems require much more data for training than rule-based or traditional statistical machine-learning systems. Hence, technological language inclusion depends to a significant extent on how much data is available, which furthers the technological gap between 'resourced' and 'under-resourced' languages. Inclusion of additional, under-resourced languages is desirable, but this becomes harder as the resources to build on are scarce. Consequently, these languages will be excluded from the use of current technologies for a long time to come and this might pose serious threats to the vitality and future active use of such languages. A useful analytical tool to assess the resources of such languages is the 'Digital Language Vitality Scale' (Soria 2017).

Advances in 'transfer learning' may help here (Nguyen & Chiang 2017; Aji et al. 2020), as well as less supervised MT (Artetxe el al. 2018). Relevant examples include HuggingFace (https://huggingface.co/Helsinki-NLP/opus-mt-mt-en) and OPUS (https://opus.nlpl.eu). There is also a need to consider the economic impact for translation companies. For example in Wales the Cymen translation company has developed and trained its own NMT within its workflow, as part of the public-private SMART partnership (https://businesswales.gov.wales/expertisewales/support-and-funding-businesses/smart-partnerships). Other companies (e.g. rws.com) have adopted similar approaches. The benefits of such technology are evident, although their use raises issues related to ownership of data, similarly to older ethical questions of who owns translation memories.

Human translators have not yet been entirely surpassed, but machines are catching up. A 2017 university study of Korean-English translation, pitting various machine translators against a human rival, came out decisively in favour of the human; but still the machines averaged around one-third accuracy (Andrew 2018). Another controlled test, comparing the accuracy of automated translation tools, concludes that "new technologies of neural and adaptive translation are not just hype, but provide substantial improvements in machine translation quality" (Lilt Labs 2017). More recently, Popel et al. (2020) demonstrated a deep learning system for machine translation of news media, which human judges assessed as ~~more accurate~~, though not yet as fluent. This was limited to news media, which is a specific linguistic register that follows fairly predictable conventions compared to conversation, personal correspondence, etc. (see Biber & Conrad, 2009); but this still shows progress.

2.1.2 Sentiment, bias

Sentiment analysis is the use of automated text analysis to detect and infer opinions, feelings, and other subjective aspects of writing - for example whether the writer was angry or happy. Extensive contributions have been made already, especially in more widely spoken languages (see Yadav & Vishwakarma 2020, for an accessible review).

Social networking sites represent a ~~continuously enriched landscape by~~ vast amounts of data daily~~, and most often hiding useful and valuable information~~. Finding and extracting the hidden "pearls" from the ocean of social media generated data constitutes one of the great advantages that sentiment analysis and opinion mining techniques can provide. Nevertheless, language spoken by social networks, like tagging, likes, the context of the comment, have yet to be explored by communities in computation, linguistics, and social sciences in order to improve the results on automatic sentiment analysis performance.

Some well known business applications include product and services reviews (Yang et al. 2020), financial markets (Carosia et al. 2020), customer relationship management (Capuano et al. 2020), ~~for~~ marketing strategies and research (Carosia et al. 2019), politics (Chauhuan et al. 2021), and in e-learning environments (Kastrati et al. 2020), among others. Most work for sentiment extraction has focused on English or other more widely used languages; and only ~~few~~ studies have identified and proposed patterns for sentiment extraction as a tool applicable for multiple languages (i.e. for bridging the gap between languages) (Abbasi et al. 2008, Vilares et al. 2017).

Focusing now on machine translation, the authors in Baccianella et al. (2010), Denecke (2008) and Esuli & Sebastiani (2006) performed sentiment classification for German texts using a multi-lingual approach. The authors translated the German texts into English language and then used SentiWordNet to assign polarity scores. Poncelas et al. (2020) discussed both advantages and drawbacks of sentiment analysis on translated texts. They reported exceptionally good results from English to languages like French and Spanish, which are relatively close to English in grammar, syntax etc.; but less good results for languages like Japanese, which are structurally more distinct.

Shalunts et al. (2016) investigated the impact of machine translation on sentiment analysis. The authors translated Russian, German and Spanish datasets into English. The experimental results showed less than 5% performance difference for sentiment analysis in English vs. non-English datasets. This gives an indication that multilingual translation can help to create multilingual corpora for sentiment analysis. ~~Alexandra et al. (2014)~~ performed machine translation to translate an English dataset of New York Times articles into German, French and Spanish using three different translators (Google, Bing & Moses). These four different texts were then used to train the multilingual sentiment classifier. For the test, the authors also used Yahoo Translator. The results supported the quality of translated text and sentiment analysis. Barriere & Balahur (2020) proposed to use automatic translation and multilingual transformer models. These are the recent advances in the NLP to solve the problem of sentiment analysis in multi-language combinations. For more detailed analysis in this area, see Lo et al. (2017).

On the issue of bias, machine learning has been applied to, for example, hyperpartisan news detection; that is, news articles biased towards a person, a party or a certain community ~~(Faerber, Qurdina & Ahmadi 2019)~~.

Bias, however, has increasingly been an issue discussed in language created automatically by machines themselves. Popular cited examples include Google Translate translating non-gendered languages like Finnish and adding gendered pronouns according to traditional gender associations: "he works, she cooks", etc. One of the challenges faced by machine learning systems and methods, in general, is judging the "fairness" of the computational model underlying those systems. Because machine learning uses real data produced by real people, to which some sort of statistical processing is applied, it is reasonable to expect that the closer those systems are to human commu-

nication, the more likely they are to reproduce all things – good and bad – about the respective population. When training corpora are skewed towards white American English-speaking males, the systems tend to be more error prone when handling speech by English-speaking females and varieties of English other than American (Hovy et al., 2017; Tatman 2017; see also https://plan-norge.no/english/girls-first; Costa-Jussà 2019). Such systems reproduce social and cultural issues and stereotypes (Nangia et al. 2020, Vanmassenhove et al. 2018).), and racial bias (Saunders et al. 2016; Lum & Isaac 2016).

Further relevant technical terminology in this field includes:

- sentiment ontologies
- enrichment and refinement
- syntactic-semantic relations
- metaphoric and implicit language properties
- sentiment evaluative terms
- multimodal contexts - for spoken data analysis performance.

Likely future developments

Work is underway to mitigate gender and other bias in machine learning, for example the automatic gendering discussed above, e.g. Sun et al. (2019), Tomalin et al. (2021). This will be especially important in light of the way automatically translated and produced texts feed into future machine learning, potentially leading to biased models exacerbating their own biases.

There are also early attempts to mobilise automated sentiment analysis for predicting suicide or self-harm, using the writing of known sufferers and victims to predict the development of conditions in others, scaled up using massive data sets (see e.g. Patil et al. 2020). From the clinical to the verificational and forensic: voice is already used as an alternative to passwords in call centres (voice signature verified by algorithm); and sentiment analysis is under development for identifying early signs of political extremist behaviour or radicalisation (see e.g. Asif et al. 2020; De Bruyn 2020).

The focus on text brings distinct limitations for other modalities - speech, sign, gesture, etc. Further studies are also required to address the cross-lingual differences and to design better sentiment classifiers. Future developments will also seek to enhance detection approaches with more accurate supervised/semi-supervised ML techniques, including transfer (transformer) models~~, are needed~~. From the linguistic standpoint, many approaches have been recently introduced, such as Google's Neural Machine Translation for delivering English text contextually similar to a certain foreign language.~~ However, machine translation does not always provide perfect accuracy when dealing with catching the sentiment of a phrase.~~

2.1.3 Text-based conversation

Within technology circles, 'chatbots' are seen as relatively primitive early predecessors to smarter and more complex successors; terms for these include ~~'dialogue systems'~~ (Klüwer 2011), "conversational interfaces" and "conversational AI". However, the term 'chatbot' has stuck and become much more common; it is therefore likely to continue dominating the popular understanding of all sorts of conversational interfaces, including dialogue systems, intelligent agents, companions and voice assistants. So we use the term 'chatbot' ~~generally~~ as an umbrella term. Current chatbots are very heterogeneous. This section is only a brief overview of all aspects of chatbot technology. For a more detailed reference see for example McTear (2020).

Chatbots embody a long-held fantasy for humanity: a machine capable of maintaining smart conversations with its creator. Chatbot technology has three principle requirements: understanding what the user said; understanding what to do next; and doing this next (usually sending a response, sometimes also performing other actions).

ELIZA (Weizenbaum 1966) is recognised to be the first chatbot. It was followed by thousands of similar machines. ELIZA was primitive: able to recognise patterns in written input, and retrieve precompiled responses. Over time, the complexity of the language comprehension capabilities increased. Audio- and video-signals were also added to the initial text-only communication.

A variety of use cases for chatbots have been explored in academic research, such as education, health, compan-

ionship, and therapy. Despite a huge amount of research, only a few of the first chatbots reached the commercial market and a wider audience (usually customer service contexts). Some car manufacturers installed conversational interfaces to be used for GPS controls and in-call hands-free phone calls. More complex, technical, forensic or clinical uses are likely some way off; indeed current early experiments have led to some alarming initial results, such as a prototype healthcare chatbot answering a dummy test patient's question "Should I kill myself?", with "I think you should" (Hutson 2021).

In 2015, social network providers realised that people use instant messengers more intensively than social networks. This was the time of the "chatbot revolution": messengers opened their APIs to developers and encouraged them to become chatbot developers by providing learning resources and free-of-charge access to developer tools. Natural Language Understanding as a service became a rapidly developing business area.

Natural Language Understanding (NLU) includes a range of technologies such as pattern-based NLU; these are powerful and successful due to a huge number of stored patterns. For instance, AIML (Artificial Intelligence Mark-up Language) forms the brain of KuKi (former Mitsuku), the Loebner prize-winner chatbot.

2.2 Speech Technology

The previous section discussed machines analysing and producing written language, including translation. The current section turns to machines working on spoken language, also including a focus on translation. Relevant terminology includes Automatic Speech Recognition (ASR) and Speech-To-Text (STT).

The human voice is impressive technology. It allows hearing people to express ideas, emotions, personality, mood, and other thoughts to other hearing people. In addition to linguistic characteristics, speech carries important para-linguistic features over and above the literal meaning of words, information about intensity, urgency, sentiment, and so on can all be conveyed in our tone, pace, pitch and other features that accompany the sounds we call words.

Think of the word 'sorry'. You could say this sincerely or sarcastically, earnestly or reluctantly, happily or sadly; you could say it in your local dialect or a more standard form; as you say it you could cry, sigh, exhale heavily, etc.; and if you heard someone saying sorry, you could immediately decode all these small but highly meaningful nuances, from voice alone. Context matters too: are you sorry only for yourself, or on behalf of someone else? Are you apologising to one person, two people, a whole country, or the entire United Federation of Planets? Fully understanding an apology means fully grasping these contextual details.

Now think about programming a machine to grasp all that, to listen like a human. It's much more than simply teaching the machine to piece together sounds into words. But progress is occurring. The evolution of speech recognition and natural language understanding have opened the way to numerous applications of voice in smart homes and ambient-assisted living, healthcare, military, education etc.

Speech technologies are considered to be one of the most promising sectors, with the global market estimated at ~~USD~~ 9.6 billion in 2020 and forecasted increase to ~~USD~~ 32.2 billion by 2027 (Research and Markets 2020). But as we have cautioned already, if private corporations are leading on these technologies, then significant concerns arise with regard to data security, privacy, and equality of access.

2.2.1 Automatic speech recognition, speech-to-text, and speech-to-speech

2.2.1.1 What is it, and how is it performed?

Automatic Speech Recognition (ASR) is the ability of devices to recognize human speech. In 1952, the first speech recognizer 'Audrey' was invented at Bell Laboratories. Since then, ASR has been rapidly developing. In the early 1970s, the US Department of Defence's Advanced Research Projects Agency funded a program involving ASR. This led to Carnegie Mellon University's 'Harpy' (1976), which could recognize over 1000 words. In the 1980s, Hidden Markov Models (HMMs) also made a big impact, allowing researchers to move beyond conventional recognition methods to statistical approaches. Accuracy, accordingly, increased. By the 1990s, products began to appear on the market. Perhaps the most well-known is Dragon Dictate (released 1990) - which, though cutting edge for its time, actually required consumer's to "train" the algorithm themselves, and to speak very slowly.

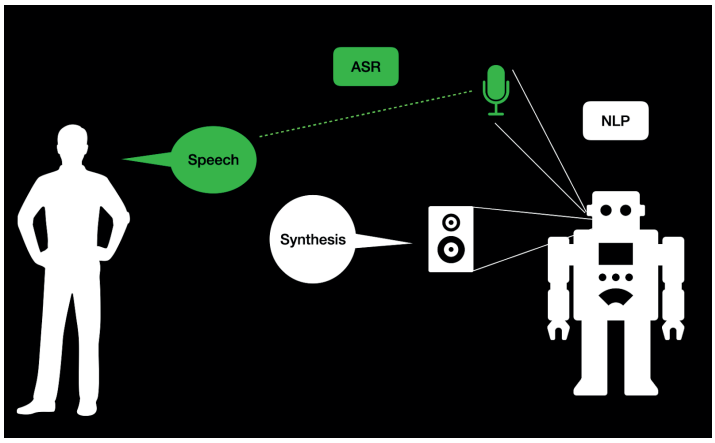


Figure 1: *Explanation goes here da da da da da da da da da da da*

Progress from this point was relatively slow until the 2010s, when Deep Neural Networks (DNNs, discussed earlier) were introduced in speech engineering.

Commercial speech recognition facilities include Microsoft Windows' inbuilt dictation facility (support.microsoft.com/en-us/fec94565-c4bd-329d-e59a-af033fa5689f), IBM Watson (ibm.com/cloud/watson-speech-to-text), Amazon Transcribe (aws.amazon.com/transcribe), and Google Speech-to-Text (https://cloud.google.com/speech-to-text).

Open source alternatives include Mozilla Deep Speech (github.com/mozilla/DeepSpeech), NVIDIA Jasper (https://nvidia.github.io/OpenSeq2Seq/html/speech-recognition/jasper), Kaldi ASR (https://github.com/kaldi-asr/kaldi), wav2letter (github.com/flashlight/wav2letter), VOSK (alphacephei.com/vosk), and Fairseq-S2T (Wang et al. 2020). Notably some of these open source facilities are developed by private companies (e.g. Facebook, NVIDIA) with their own incentives to contribute to other products in their portfolio.

A recently founded EU-funded project, MateSUB (matesub.com), is leveraging these kinds of capabilities specifically for adding subtitles following speech recognition. Machine translation of subtitles was the topic of SUMAT project http://www.fp7-sumat-project.eu/.

2.2.1.2 What are some of the challenges?

Many challenges remain for machines to faithfully and reliably decode human speech. These include at least the following:

- Different words that sound the same, like 'here'/'hear', 'bare'/'bear'. These are known as 'homophones' (i.e. same sound) and require more than just the sound alone to understand.
- Rapidly switching between dialects or languages ('code-switching'), which is extremely common in normal human conversation around the world
- Variability in the volume or quality of someone's voice, including things like illness, or physical blockages like chewing food
- Ambient sounds like echoes or road noise
- Transfer influences from one's first language(s) to second languages (Elfeky et al. 2018, Li et al. 2017).
- All sorts of other conversational devices we use, like elisions (skipping sounds within words to say them more easily), or repair (making a small error and going back to correct it)
- Paralinguistic features: pace, tone, intonation, volume

For all these various levels of meaning and nuances of speech, there are relatively few annotated 'training sets', that is, databases of speech that contain not only transcribed speech but all that other necessary information, in a format a machine could understand. This is especially acute for lesser-resourced languages. And if systems are

initially trained on English (~~and or~~ English paired with other languages), and then transferred to other languages and language pairs, there could be a bias towards the norms of the English language, which might differ with other languages.

The issue of speaker variation (individual variation, non-standard dialects, learner varieties, etc.) requires greater attention. Equal access to speech recognition technology will depend heavily on this. A standardized framework for describing these is under development in ISO (International Organization for Standardization).

Likely future improvements

Traditional speech technologies require massive amounts of transcribed speech and expert knowledge. While this works fairly well for 'major' languages, the majority of the world's languages lack such resources. A large amount of research is therefore dedicated to the development of speech technologies for 'low-resource' or 'zero-resource' languages. The idea is to mimic the way infants learn to speak, spontaneously, directly from raw sensory input, with minimal or no supervision. A huge step towards unsupervised speech recognition was made when Facebook released wav2vec (Schneider et al. 2019) and its successor wav2vec 2.0 (Baevski et al. 2020), which is able to achieve a 5.2% word error rate using only 10 minutes of transcribed speech. It learns speech representations directly from raw speech signals without any annotations, requiring no domain knowledge, while the model is fine-tuned using only a minimal amount of transcribed speech. This holds great promise, though success will depend on factors including accessibility, privacy concerns, and end user cost.

Mozilla Deep Speech, accompanied with its annotated data CommonVoice initiative (see next section), aims to employ transfer learning. That is, models previously trained on existing large annotated corpora, such as for American English, are adapted and re-trained with smaller annotated corpora for a new domain and or language. Such an approach has been proven to be viable for bootstrapping speech recognition in a voice assistant for a low-resourced language. Companies like Google, as well as many university AI research groups, are busily attempting to apply self-supervised learning techniques to the automatic discovery and learning of representations in speech. With little or no need for an annotated corpus, self-supervised learning has the potential to provide speech technology to a very wide diversity of languages and varieties: see for example https://icml-sas.gitlab.io/

Further challenges ahead for automated subtitling include improved quality, and less reliance on human post-editing (Matusov et al. 2019).

2.2.2 Voice Synthesis

2.2.2.1 What is it, and how is it performed?

Voice synthesis is the ability of machines to produce speech sounds artificially.

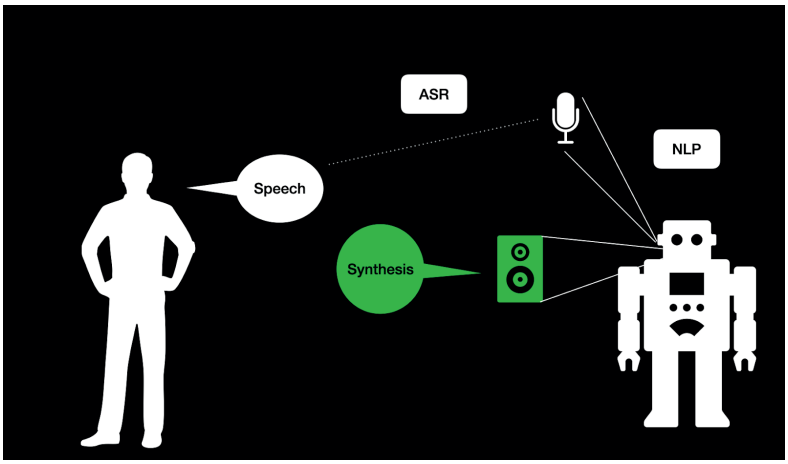


Figure 1: *Explanation goes here da da da da da da da da da da da*

With origins as a niche research field - restricted to the most highly trained specialists - speech synthesis is now a large domain with people of varying specialisms producing core components of successful commercial products. The success of voices like Amazon's Alexa and Apple's Siri were built on years of work on speech modelling and parametrization. Contemporary systems, based on advanced neural modelling techniques, get us closer to bridging the quality and naturalness gap while still offering flexibility and control. They are capable of modelling challenging heterogeneous data, i.e. data that contains multiple sources of variation such as speakers and languages, non-ideal recording conditions, and expressive and spontaneous speech.

The current cutting edge in voice synthesis is to go beyond simply creating a 'life-like' robot voice, and instead fully mimicking a real person. This can be achieved by mobilising the recently discussed 'deep learning' advances in AI. A voice synthesis algorithm can take a recording of a person's voice (ideally a relatively long one with a range of different sounds), and then apply deep learning techniques to assemble building blocks of sounds in that person's voice - their accent, their pitch, tone, pace, and use of pauses and fillers. The machine can then create entirely new words, in new combinations, with that person's unique cadence, and with smooth transitions between words.

There is a vibrant and jostling market of companies offering automated voice mimicry, for example ReSpeecher (https://www.respeecher.com/). They offer revoicing of a user's voice of another person, and automated dubbing of video into other languages in the actors' original voices. MateDUB (matedub.com) is an EU-funded project for automatically creating dubbed audio. Anyone is free to upload their voice and set a price for using the new automated voice. This means it isn't free, but it is significantly cheaper than voice actors recording everything. Another service is https://videodubber.com/. Amazon is also working in this area: https://arxiv.org/pdf/2001.06785.pdf.

Mozilla's Common Voice (https://commonvoice.mozilla.org) aims to bring some of these capabilities to the world in a fully free and open-source way. This is intended to complement Mozilla's 'Deep Speech' speech recognition utility mentioned earlier in this report.

Further relevant terminology in the field of voice synthesis includes:

- Text processing and signal processing, including text normalisation, letter-to-sound conversion, short term analysis of sequential signals, frequency analysis and pitch extraction
- Concatenative speech synthesis, including diphone synthesis and unit selection synthesis
- Statistical parametric based speech synthesis, including parametrizing speech using vocoders and acoustic modelling using HMMs
- Currently: deep neural networks for acoustic modelling and waveform generation (replacing decision tree HMM-based models and vocoders)
- Advanced techniques for acoustic modelling using sequence-to-sequence ('seq2seq') (Hewitt & Kriz 2018) models for end-to-end ('e2e') speech synthesis (Taylor & Richmond 2020).

2.2.2.2 What are some of the challenges?

Some challenges are similar to voice recognition as noted earlier, for example the availability of data to train machines: Mozilla Common Voice for example requires 10,000 hours of speech to add a new language. Work is ongoing to bridge this gap by applying 'transfer learning', as discussed earlier; see for example Jones (2020) on development of a voice assistant for Welsh. Other common challenges with voice recognition include linguistic structural issues like homophones and code-switching.

Challenges specific to voice synthesis include:

- Achieving "naturalness", that is, passing for human according to human test subjects;
- Persuasiveness and trustworthiness of automated voices; for example Dubiel et al., 2020, compare the persuasiveness of a chatbot speaking in different *styles*: "debating style vs. speech from audio-books", while Gálvez et al. (2017) discuss variability in pitch, intensity and speech rate linked to judgements of truthfulness;
- measures to detect and intercept malicious imitation used for identity fraud.

Speech recognition and voice synthesis can be combined together in various combined software applications. The application of machine translation to spoken language is more recent, though gaining rapidly in use. The principles

do not differ significantly. The technology that enables two or more interlocutors to communicate in their own language in near real time already exists and is available to common users. Jibbigo, which was developed at Carnegie Mellon University in the early 2000s, is a good example. It started as an iPhone app to provide speech-to-speech translation between English and Spanish, but later included more languages, and was also ported to Android. In addition to popular tools like Microsoft Translator and Google Translate, a number of services are available that provide users with voice translation, or allow them to translate text and then read it aloud: iTranslate, Speak to Voice Translator, VoiceTra, Voice Translator 2020, Translate All: Translation Voice Text & Dictionary, Speak & Translate, SayHi Translate and Naver Papago – AI Translator.

Skype has also developed its own voice translation system, Skype Translator, which enables translation of text from/into over 60 languages, and speech from/into 11 languages (or, to be more precise, language varieties), including Chinese (Simplified and Traditional), English (UK and US), French, German, Italian, Japanese, Portuguese, Russian and Spanish. More languages will certainly follow in the near future.

The development of applications for machine translation of haptic languages has been slower, especially given the limitations underlying haptic language technologies. Yet, systems have evolved significantly in recent years. RoboBraille.org is an online service designed to translate text into braille, rendered as either six-dot or eight-dot Braille (Christensen 2009). A similar functionality is provided by BrailleTranslator.org. The Dot Translation Engine and its related tactile device, the Dot Mini, allows the visually impaired to translate documents rapidly into Braille (https://hyperinteractive.de/portfolio/dot-translation-engine/). For a current overview of technologies taking automated approaches to Braille, see Shokat et al. (2020). The Concept Coding Framework (CCF) is an attempt to provide a generic approach to make content interoperable in any language/communication modality: conceptcoding.org.

Most machine translation of speech currently works by interpreting voice patterns into words, then assembling this into speech, before 'reading out' that text in a synthesised voice. Further advances in machine translation are working towards direct speech-to-speech translation, for example the Google Translatotron: https://google-research.github.io/lingvo-lab/translatotron/.

2.3 Visual and tactile elements of interaction

We previously discussed sign language and the contribution of factors like facial expression and body posture to the meaning of sign. Sign languages make use of a wide range of very specific bodily cues. In spoken language too, the body is used, though less precisely. For hearing people, the body is a wide paint roller; for signers, it is a fine brush. As we also noted earlier, machine analysis of the visual elements of interaction has quite some way to go in comparison to voice and text.

2.3.1 Facial expression, gesture, sign language

Sorgini et al. (2018) give a review of progress in this area. Small-sized, tailor-made, low-cost haptic interfaces are in development. These interfaces will be integrated with common devices such as smartphones, contributing to a massification of sensory assistants among those impaired. This will also mean a move from invasive sensory implants to less invasive alternatives (ibid.).

Machines have different tasks ahead of them in gauging the importance of body movements. For example, there is work aiming to simply identify who is speaking based on gesture (Gebre & Heskes, 2013). The task gets more and more specific until we get down to the much more precise work of interpreting sign language, which is still some appreciable way from being within the command of machines. As explained by Jantunen et al. (2021), research into automated sign language detection, processing and translation is important and worthy; but currently no automated systems are anywhere close to full functionality.

First of all, as we noted earlier in this report, sign languages are not just a visual rendering of spoken language. They are entirely separate languages, with their own grammar. These grammars have attracted significant research attention. A relevant source for comparative grammars of sign languages, potentially relevant for computational purposes, is the Sign-Hub project (https://sign-hub.eu/). Additionally, rule-based machine translation has recently demonstrated promising results for sign language, given how it represents grammar (Filhol et al. 2016). But these

endeavours have faced serious limitations. Sign-hub has adopted the formal (generative) theory as its starting point, so the blueprint of sign language grammars is based on the study of spoken language, and especially English. The sensor gloves, too, require further development, not the least because the data used in the test/evaluation of the sensor gloves was only ten finger alphabets and number signs plus one emblem gesture, "I love you". That is eleven sign types, and only in American Sign Language. Consequently, it still remains to be established how the information produced as part of the Sign-hub can be used to contribute to a robust and generally accepted SL translator. Hadjadj et al. (2018) proposed an alternative approach to the grammar of French Sign language that takes into account the additional grammatical characteristics of sign language. There is much progress left to make here.

One problem for machine learning of sign language is shared with minority spoken languages: scarce and unstructured data. For spoken languages, this is somewhat easier to solve than for sign: just feed more data into the same systems used for larger languages; and/or improve 'transfer learning' approaches which we discussed earlier. For sign, the problem is much more complex. Systems for automatic recognition of speech (audio only) and writing simply cannot understand sign. Sign language corpora are not only smaller than spoken corpora; they are much harder to gather. Remember that sign is a different modality to speech; sign corpora must be annotated manually for machines to learn from, which is demanding and very time-consuming (Jantunen et al. 2021). The main enigma for machine translation of sign language in the near future is the bounty of unconventional characteristics that exist in all sign languages – the so-called indicating and depicting aspects of signed utterances - which are not easily translatable.

As a result, data sets are fewer, and those that exist have been collected under "controlled environments with limited vocabulary" (Camgöz et al. 2018: 7785). Eventually, machines learn from annotated models, and not directly from video as would be required to capture the inherently multimodal nature of sign. Attempts have been made to make progress in this area, by finding new ways to collect multimodal data (Camgöz et al. 2018) and to program intelligent avatars to produce sign language after translating speech (Stoll et al. 2018). Neural networks generate video content without relying on extensive motion capture data and complex animation. Nevertheless, as the authors caution, this work is rare and foundational, and still far behind the progress achieved so far by research into writing and speech. Another issue is the confidentiality of the individuals involved in the production of sign language samples collected for shared datasets: as a recent study suggests, signers can be recognized based on motion capture information (Bigand et al. 2020).

Although all progress is welcome, the technological advances in the field of sign language have been slow, when compared to the pace at which written and spoken language technologies have evolved.

As Jantunen et al. (2021) predict, machine translation of sign languages will face at least three important challenges for some time to come (that is, long after similar obstacles are overcome for spoken and written modalities): (a) multimodality, wherein meaning is made not only with hand gestures but also with a rich mix of gesture, facial expression, body posture, and other physical cues, yet even the most advanced detection systems in development are focused only on hand movement; (b) there are hundreds to thousands of sign languages, but research so far has focused on 'major' sign languages, so the complexity of sign language communication and translation is higher than gesture-recognition systems currently take into account; (c) meaning often also depends on socio-cultural context and signers' knowledge of each other's lives, which machines cannot know (and training them to find out provokes major privacy concerns).

In section 1.3.3 we discussed sign language. In section 2.3.1 we expand sign recognition and synthesis - including its many persistent limitations.

2.3.2 Tactile expression and haptic technology

Haptic assistive technologies build upon tactile sense to offer sensory information to deaf, blind and deaf-blind individuals when communicating with the non-disabled community. The visual and auditory cues received by the machine are converted into haptic feedback; that is, targeted pressure on the skin in a pattern that corresponds to meaning.

It is a common misconception that the human-machine interaction of blind people requires a special, highly specific and sophisticated, Braille-enabled computer. Currently, a blind person can use an ordinary computer, equipped with an ordinary keyboard; no physical adaptations or alterations are required. They touch-type, perhaps using automated audio readouts of each keystroke to confirm. Instead of a mouse, shortcut keys are used (many

such shortcuts pre-date the invention of the computer mouse and are quite standard). The information shown in the computer screen is vocalised by a "screen reader" using voice synthesis methods discussed earlier. Screen readers also allow users to control a Braille terminal - a device connected to the computer to show the computer screen in Braille. Information can also be printed using a Braille printer.

Research into so-called tactile sign language is even scarcer (Willoughby et al. 2018), partly because tactile sign languages are still developing stable conventions. The group of deaf-blind signers is highly heterogeneous (ibid.), and the influence of sociolinguistic or fluency factors (e.g. at what life stage did the tactile sign language acquisition occur) is still unknown. Because so much pragmatic information in visual sign languages is communicated non-manually, such meaning can hardly be made - if at all - with tactile signs. We are at our most cautious when discussing progress in this area.

2.4 Pragmatics: the social life of words

As well as words and body movements, machines will also need to understand the *purpose* of each utterance, and how it changes the world around us. This brings us into the realm of *pragmatics*, including the subtle negotiation of politeness, sincerity, honesty, deception, and so on. Pragmatics has a long history as an academic discipline, and more recently the interdisciplinary field of Computational Pragmatics has arisen, somewhat as a subdiscipline of Computational Linguistics.

Computational Pragmatics is perhaps less developed than other areas of Computational Linguistics, basically due to two main limitations: the need to further develop and structure Pragmatics itself as a theoretical field; and the need to further develop other subdisciplinary areas of Computational Linguistics, not least gesture recognition.

The first limitation is the bounds of pragmatics itself. This should not be underestimated. It remains quite an enigma how we combine such a rich and diverse range of actions to achieve things in conversation. There is simply a vast and splaying prism of pragmatic meaning-making, the myriad social meanings and intentions that go into what we say, and the similarly disparate array of effects these have in the world. Pragmatics is simply enormous, and very far from reaching any kind of unifying theories - it is sometimes affectionately (or exasperatedly) labelled "the pragmatics wastebasket" (Yule 1996: 6) for accommodating more or less all communicative phenomena that do not neatly fit in other linguistic levels (syntax, morphology, etc.). To elaborate a little further, the myriad phenomena at play include at least:

- how we encode physical distance ("You're too far away", "Come closer");
- how we select the right terms to address each other ("Madam", "buddy", "Mx", etc.);
- tacit cultural knowledge, or prior understanding that directly affects word choice, like who I am and who you are, where we are, what will happen if you drop a ball vs. a glass, etc.;
- how we convince people to do things by appearing authoritative, weak, apologetic, etc.;
- discourse markers and fillers ("hmm", "uhuh", "right").

And for every single one of these, there is bountiful and wonderful but baffling and unknowable diversity and change - across languages and cultures, within smaller groups, between individuals, and in the same individual when talking to different people at different times. All this adds up to quite a challenge for machines to understand pragmatic meaning.

The second limitation noted above is the need to further develop other fields of Computational Linguistics. Recognising and classifying pragmatic phenomena first relies on recognising and classifying other linguistic phenomena (e.g., phonological, prosodic, morphological, syntactic or semantic). If someone tells a friend they are short of money, it could imply a request, a dispensation, a simple plea for pity, or something else; a machine cannot know which without first knowing about different ways of describing money, poverty, and so on; as well as the subtle but vital combination of gestures, facial expressions and other movements that could contribute to any of these intended meanings. All this might entail the incorporation into pragmatics-related corpora of sound at a larger scale and its processing and annotation with adequate schemes and formats.

Archer et al. (2008) mention two definitions of computational pragmatics: "the computational study of the relation between utterances and action" (Jurafsky 2004: 578); and "getting natural language processing systems to reason in

a way that allows machines to interpret utterances in context" (McEnery 1995: 12). As far as pragmatic annotation is concerned, it is noted that "the majority of the better-known (corpus-based) pragmatic annotation schemes are devoted to one aspect of inference: the identification of speech/dialogue acts" (Archer et al. 2008: 620).

Some projects developed subsequently - such as the Penn Discourse Treebank (https://www.cis.upenn.edu/~pdtb/) - have also worked extensively in the annotation of discourse connectives, discourse relations and discourse structure in many languages (see e.g. Ramesh et al. 2012; Lee et al. 2016; Webber et al. 2012).

Progress in the last two decades includes attempts to standardise subareas of Pragmatics, such as discourse structure (ISO/TS 24617-5:2014), discourse relations (ISO 24617-8:2016) speech act annotation (ISO 24617-2:2020), dialogue acts (ISO 24617-2:2020), and semantic relations in discourse (ISO 24617-8:2016); and even to structure the whole field of Computational Pragmatics and pragmatic annotation (Pareja-Lora and Aguado de Cea 2010; Pareja-Lora 2014) and integrate it with other levels of Computational Linguistics and linguistic annotation (Pareja-Lora 2012). Further current research concerns, for instance, the polarity of speech acts, that is, in their classification as neutral, face-saving or face-threatening acts (Naderi & Hirst 2018).

However, as Archer, Culpeper & Davies (2008) indicate, "[u]nlike the computational studies concerning speech act interpretation, [...] corpus-based schemes are, in the main, applied manually, and schemes that are semi-automatic tend to be limited to specific domains" (e.g., "task-oriented telephone dialogues"). This is only one of the manifold limitations of research in this area.

All this could be solved, to some extent, by suitable annotated gold standards to help train machine learning tools for the (semi-)automatic annotation and/recognition of pragmatic phenomena. These gold standards would need to include and integrate annotations pertaining to all linguistic levels - as discussed above, a machine may struggle to identify pragmatic values if it cannot first identify other linguistic features (for instance politeness encoded in honorifics, pronouns, and verb forms). These annotated gold standards would be quite useful also for the evaluation of any other kinds of systems classifying and/or predicting some particular pragmatic phenomenon.

Another big limitation in this field, as discussed above, is about the journey our words take out there in the real world. Much of our discussion so far in this report is all about the basic message we transmit and receive: words, signs, and so on. But human language is much more complex. The basic message - the combination of sounds, the array of signs and gestures, that make up our 'utterances' - are absolutely not the end of the story for language. When we put together the words 'Let me go', those words have a linguistic meaning; they also carry an intention; and then subsequently (we hope) they have an actual effect on our lives. These are different aspects of our language, all essential for any kind of full understanding. Neurotypical adults understand all these intuitively but they must be learned; and so a machine must be trained accordingly. There have been some advances but constraints remain (mentioned below and also pervasively in this document), for example:

1. Higher-order logical representation of language, discourse and statement meaning are still partial, incomplete and/or under development.
2. Perhaps also as a consequence, computational inference over higher-order logic(s) for language (e.g., to deal with presuppositions or inference) require further research to overcome their own current limitations and problems. "Indeed, inference is said to pose "four core inferential problems" for the computational community: abduction [...], reference resolution [...], the interpretation and generation of speech acts [...], and the interpretation and generation of discourse structure and coherence relations [...]" (Archer, Culpeper & Davies, 2008). The first of these, abduction, means roughly inference towards the best possible explanation, and has proved the most challenging for machines to learn; no great progress is expected here in the next decade. But progress on speech acts and discourse structure (and/or relations) has been robust for some widely-spoken languages; and some resources and efforts are being devoted to the reference resolution problem (that is, reference, inference, and ways of referring to physical space), in the fields of (i) named entity recognition and annotation and (ii) anaphora (and co-reference) resolution.

The final big limitation of this field is the strong dependency of pragmatic features on culture and cultural differences. Indeed, once identified, the values of these pragmatic features must be interpreted (or generated) according to their particular cultural and societal (not only linguistic) context. That pragmatic disambiguation is often a challenge for humans, let alone machines. Take 'face-saving' or 'face-threatening' acts: for example we attempt face-saving for a friend when we say something has gone missing, not that our friend lost it; while face-threatening acts, by contrast, are less forgiving.

Interpretation or formulation of face-saving and face-threatening acts are highly culture-dependent. This also affects, for example, the interpretation and production of distance-related features (any kind of distance: spatial, social, temporal, etc.). Earlier we mentioned some levels of pragmatic meaning - our basic literal message, our intention, its possible effects in the world. Understanding face-saving is a key part of managing those things, and they all differ according to who we are talking to. It is almost impossible to understand pragmatic meaning without understanding a huge amount of overlapping social information. Nothing is understood before everything is understood.

In the end, all these aspects entail the codification and management of lots of common (or world) knowledge, information, features and values. Machines might be more able to process all these items now than in the past by means of big data processes and techniques (such as supercomputation or cloud computing). However, all these items still need to be identified and encoded in a suitable computer-readable format. The community of linked open data is working hard in this aspect and their advances might help solve this issue in due course (Pareja-Lora et al. 2020).

Likely future developments

Seemingly, the likely future developments in this field might be the application of all this progress to the areas of:

1. Human-machine interaction (e.g., chatbots). As above, chatbots can decode the sounds into words and grammatical structures, but they are much less adept at understanding what we want to achieve socially with our words, much less the varied interpretations of our intentions by those around us. Chatbots therefore make lots of mistakes and the human user usually feels frustrated. Efforts will be directed towards these issues.
2. Virtual reality avatars. More and more sociolinguistic knowledge will be incorporated in the programming of these entities, to make them increasingly natural and user-friendly.
3. Machine translation and interpretation. Machine (or automatic) interpretation is still a very young field, since it has to somehow encompass and integrate both natural language processing and generation. Thus, it needs both of these areas to progress before it can further be developed. However, it seems that the time is ripe for a major leap forward in this field, and machine interpretation should blossom in the coming years, hand in hand with the advances within Computational Pragmatics.

2.5 Politeness

Linguistic politeness concerns the way we negotiate relationships with language. This is in some senses a sub-disciplinary area of pragmatics. We design our speech in order to further our intentions. For that, we pay attention to 'face'. In common parlance, to 'save face' is to say something in a way that minimises the imposition or embarrassment it might cause. It is simple everyday conversational diplomacy. Politeness theory builds on this simple insight to interrogate the various ways we attend to other people's 'face needs', their self-esteem and their sense of worth. Neurotypical adults have an intuitive sense of interlocutors' 'face needs'. That sense enables a choice about whether we want to either uphold or undermine those needs. Do we say 'I'm sorry I wasn't clear' or 'You idiot, you completely misunderstood me!', or something in between these extremes? They 'mean' the same thing, but they attend to face needs very differently.

How we attend to face needs will depend on the nature of the relationship, and what we want to achieve in interaction. There is the basic classical distinction between 'positive face' (the desire to be liked) and 'negative face' (the desire not to impose on people). Brown and Levinson (1987) is a commonly used foundational text in the discipline, though obviously somewhat dated now. Since then, the study of politeness has proceeded to explore various aspects of extra-linguistic behaviour and meaning. This has resulted in increased insight, but has somewhat fragmented the field away from the kinds of unifying, standardised principles represented by Brown and Levinson (1987). But developers of machine learning systems need a little more stability; and so, current studies in machine learning of politeness are in the slightly curious position of continuing to apply categories from Brown and Levinson (1987) to machine learning - for example Li et al. (2020) on social media posts in the US and China; Naderi & Hirst (2018) on a corpus from the official Canadian parliamentary proceedings; and Lee et al. (2021) on interactions between robots and children. All these studies use categories from Brown & Levinson (1987) as a basis for machine learning. The research teams themselves are not failing per se in their background reading;

rather, this reliance on older work simply reflects the relatively early stage of the technology, and the need for tight structures and categories from which machines can learn. This in turn indicates the distance left to go in designing pragmatically aware machines.

The difficulty (for machines and indeed often for humans) lies in the heavily contextual nature of linguistic politeness within one language and across languages, and within individuals and different social groups. Culturally aware robots will need to understand that some sentiments will be expressed in some communities, while in others it is not acceptable to express them. Some verbal reactions may be acceptable in some cultures, in others less so or not at all. This will be important for social inclusion and justice in multicultural societies.

The above reports are also based on text alone; or if they cover gesture, they focus on production of politeness gestures by machines, not interpretation of human gestures by machines. As discussed above in relation to sign language, this is simply beyond the purview of the current state of the art. Bots can themselves approximate polite behaviour, but cannot distinguish it in users.

Likely future developments

The more naturally humans communicate with intelligent machines, the more important the role that politeness will play in both directions. Currently, communication is mainly targeted at achieving quite specific tasks; in these contexts, politeness may not be necessary, may even be superfluous and over-complicate the interaction. "Alexa, play my favourite song" rather than "Alexa, could you please play my favourite song?". Looking ahead, as machines improve their pragmatic awareness, this will most probably result in a change towards more politeness in human-machine communication. At the same time, humans will also expect machines to address them with the honorific structures (e.g. formal French *vous* vs. informal French *tu*) that correspond to the level of formality and mutual familiarity with each other in a way that parallels human interactions with friends or strangers.

A forward-focused review of useful pragmatic principles for future chatbot design is provided by Dippold et al. (2020).

3 Gadgets & gizmos: human-integrated devices

Summary and overview

Two major technological fields will influence language in the human-machine era: Augmented Reality (AR), and Virtual Reality (VR). Both of these are currently somewhat niche areas of consumer technology, but both are subject to enormous levels of corporate investment explicitly targeting wide adoption in the coming years. By 2025, the global market for AR is forecast to be around $200 billion, and for VR around $140 billion.

AR is currently widely available: embedded into mobile phones in apps like Google Translate which can translate language in the phone's view; and in dedicated headsets and glasses that overlay the user's visual field with richer and more detailed information. Progress in AR is rapid, though to date the headsets and glasses have been mostly used in industrial settings due to their somewhat awkward and unsightly dimensions. In the next year or two, AR tech will shrink down into normal sized glasses, at which point the above-noted market expansion will begin in earnest.

AR eyepieces will combine with intelligent earpieces for more immersive augmentation. These devices will enable other AI technologies to combine and transform language use in specific ways. Two in particular are augmentation of voice, and augmentation of facial/mouth movements. As covered in the previous section, advances in speech recognition and voice synthesis

are enabling automatic transcription of human language, rapid translation, and synthetic speech closely mimicking human voices. These will be straightforward to include in AR eyepieces and earpieces. That in turn will enable normal conversation to be subtitled for clarity or recording, amplified, filtered in noisy environments, and translated in real time. Meanwhile advances in facial recognition and augmentation are beginning to enable real-time alterations to people's facial movements. At present this is targeted at automatically lip-syncing translations of video, making the mouth look as though it is producing the target language as well as speaking it. Embedded into AR eyepieces, this will also allow us to talk to someone who speaks another language, hear them speaking our language (in their own voice) while their mouth appears to make the translated words. This is what we mean by speaking through technology. The devices and algorithms will be an active contributor to our language, tightly woven into the act of speaking and listening.

VR is similarly the site of rapid technological progress, and vast corporate investment. Devices currently on the market enable highly immersive gaming and interaction between humans. Future advances in VR point to much more transformative experiences with regard to language. Mostly this relates to the combination of VR with the same technologies mentioned earlier, plus some extra ones. Augmentation of voice and face, noted above, will enable much more immersive first-person and multi-player games and other interaction scenarios, including translation and revoicing. The main distinguishing feature of VR will be in the addition of future improved chatbot technology.

Currently chatbots are somewhat limited in their breadth of topics and complexity of responses. This is rapidly changing. In the near future chatbots will be able to hold much more complex and diverse conversations, switching between topics, registers, and languages at pace. Embedded into virtual worlds, this will enable us to interact with highly lifelike virtual characters. These could become effective teachers, debaters, counsellors, friends, and language learning partners, among others. The potential for language learning is perhaps one of the most transformative aspects. VR is already used for language learning, mostly as a venue, with some limited chatbot-based interaction. This will change as VR chatbots in virtual characters evolve. The possibility of learning from endlessly patient conversation partners, who would never tire of you repeating the same word, phrase or sentence, who would happily repeat an interaction again and again for your confidence, could truly upend much about pedagogical practice and learner motivation and engagement.

AR too will enable language learning in different ways. As we noted above, it will soon be possible to translate in-person conversations in AR earpieces. AR eyepieces could also show you the same overlain information in whatever language it supports. There are tools already available for this, and their sophistication and number will grow rapidly.

Next, law and order. Machines are already used for some legal tasks, and recent years have seen marked growth in the market for 'robot lawyers' able to file relatively simple legal claims, for example appealing parking tickets, drafting basic contracts, and so on. This too is set to grow in breadth and complexity with advances in AI, moving into drafting of legislation, and regulatory compliance.

Somewhat relatedly, machine learning will increasingly take on enforcement and forensic applications. Machine learning is already applied for example to plagiarism detection in education, to personality profiling in human resource management, and other comparable tasks, as well as identifying faces in crowds and voices in audio recordings. Illicit uses include creation of fake news and other disinformation, as well as impersonation of others to bypass voice-based authentications systems. All these are likely to expand in due course, especially as these capabilities become embedded into wearable devices.

In health and care, language AI will enable better diagnosis and monitoring of health conditions. Examples include diagnosis of conditions that come with tell-tale changes to the patient's voice, like Parkinson's or Alzheimer's disease. AI can detect these changes much earlier than friends or relatives, since it is not distracted by familiarity or slow progression. Ubiquitous 'always on' listening devices will be primely placed for this kind of diagnosis and monitoring.

For sign language, certain devices are being developed for automatic recognition and synthesis of sign. These have tended to focus exclusively on detecting the handshapes made during signing, but signing actually involves much more than this. Fully understanding and producing sign also involves a range of other visible features, including gesture, gaze, facial expression, body posture, and social context. These are currently out of reach for current and even planned future technologies. For this reason the Deaf community will benefit much more slowly from technological advances in the human-machine era.

Moreover, as we noted previously, and as we will continue to note in this report, all these exciting advances will not work equally well for everyone. They will obviously be unaffordable for many, at least in the early stages. They are also likely to work better in larger languages than in smaller and less well-resourced languages. Sign language have the additional obstacles just noted.

AR has additional major implications for a range of language professionals - writers, editors, translators, journalists, and others. As machines become more able to auto-complete our sentences, remember common sentences or responses, and indeed generate significant amounts of content independently, so too the world of language work will change significantly. There will still be a role for humans for the foreseeable future, but language professionals, perhaps more than anyone, will be writing and talking through technology.

All this will also in turn provoke major dilemmas and debates about privacy, security, regulation and safeguarding. These will come to the fore alongside the rise of these technologies, and spark into civic debate in the years to come.

3.1 Augmented Reality

In recent years, research has demonstrated that it is technically possible to augment our reality with digital information in many different ways. This is thanks to advanced visual and audio tracking and registration algorithms. The first mediums for this will be smartphones and wearables such as glasses; but other small projection devices in our surroundings could project virtual displays, either on to surfaces or as holograms. Contexts could include homes and offices, streets in smart cities, or even micro-miniaturised devices on our arms or legs, using our own body as a projection screen and interface. The possibilities are diverse, and they all point towards a future of seeing extra information about the world projected in our eyes and ears.

3.1.2 The visual overlay

The devices we review here currently exist at prototype stage or are used in specialist commercial contexts. As a first example, Google Glass was famously a consumer flop (too bulky and awkward looking), but it nevertheless found great success in enterprise settings, flashing up details of products and parts into workers' visual fields. One such deployment (among many worldwide) is a mobile automotive repair service in Argentina, deploying Google Glass to deliver real-time supervision of its mechanics:

Figure X: http://www.igs.com.ar/portfolio-item/fca-fiat-remote-support-with-google-glass/

This is useful and used in a range of other professional contexts, but the form factor is still slightly too bulky for streetwise social contexts. Nevertheless, the technology is being rapidly refined and miniaturised to become inconspicuous and sleek. There is currently much media hype surrounding Apple Glass, rumoured for release in 2022. Apple is being characteristically secretive about this; but certain competitors are saying more, for example Facebook's 'Project Aria': https://about.fb.com/realitylabs/projectaria/. Other notable rivals include Nreal (nreal.ai), and MagicLeap (magicleap.com). With all these advances in view, the global market for AR devices is forecast to rise from $3.5 billion in 2017 to around $200 billion by 2025 (Statista 2020).

Altogether, advances in AR will combine for a comprehensive and varied means to deliver visual augmentations. Now, the challenge is to take advantage of such overlay to bring applications that go beyond mere visual augmentations, and exploit the multimodal nature of human interactions. In this respect, combining visual tracking and augmentations with language interactions would be particularly interesting and symbiotic. A major milestone, and a cornerstone of the human-machine era, will be feeding information into conversation in real time.

3.1.3 Augmenting the voices we hear

The visual overlay we just discussed lends itself very clearly to a nascent capability we discussed earlier in the report: adding live subtitles to spoken language. This will be a relatively simple port of one technology onto another, and will be one of the first augmentations of spoken language. The key difference will be seeing live subtitles 'in the wild', in live interaction, projected into your visual field.

Two technologies will come together to improve and refine recognition of spoken words. The first is improved audio recording (and noise cancellation). The second is visual recognition of 'visemes', facial movements associated with particular words, potentially enabling lip-reading for use in high-noise environments, or over distances. There is ongoing incremental progress in recognising visemes (e.g. Peymanfard et al. 2021). Recognition of sign languages could also be embedded, if and when that improves - although as we noted earlier, that will lag behind significantly, to the disadvantage of the Deaf community. Live subtitling of speech will help to a degree, but for many Deaf people written language is a second language, less readily accessible than sign.

These two technologies (speech recognition and viseme recognition) are under development as we discussed earlier; and both will enable enhancements of eyewear and earwear. This will allow us to talk in very noisy environments, or even in the dark by using night vision. Meanwhile, for hearing people, future earpieces will incorporate improved automated translation, in order to deliver live translated audio of people speaking other languages.

Furthermore, advances in voice synthesis will enable the earpiece to mimic the voice of whoever you are talking to. Earlier we discussed the current market of voice synthesis companies, including ReSpeecher. Such companies have diverse future plans, eyeing the unification of this software with new hardware and integration in different life contexts. ReSpeecher's own plans include: "people who suffer from strokes, cancer, ALS, or other ailments will use Respeecher to replicate their own voices and speak naturally". This instantaneous revoicing could quickly progress beyond clinical settings and be combined with real-time automated translation, for everyday use; you could hear people translated into your language, in their own voice, or indeed in your voice, or any voice you choose.

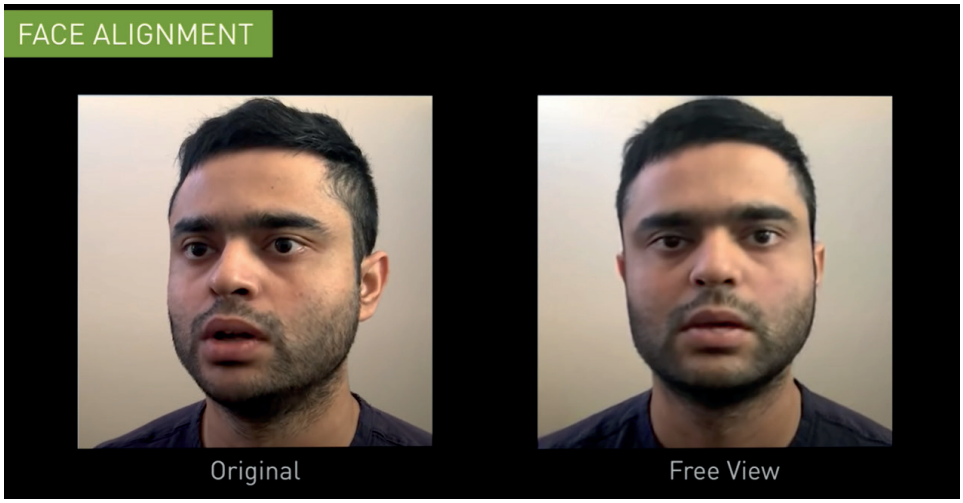


Figure X: source: https://youtu.be/eFK7Iy8enqM

3.1.4 Augmenting the faces we see

Advances in algorithms and techniques for morphing and blending faces are also opening a different sort of augmentation. To begin with video conferencing - which has increased exponentially during the COVID-19 lock-downs - participants will often look towards their screen and away from the webcam, sometimes in an entirely different direction. Efforts are underway to use deep learning to learn the shape of the user's face, and reanimate it in the webcam video so that they appear to be looking directly at the camera at all times. In 2020, NVIDIA announced early developer access to its new 'Maxine' platform. Designed to streamline various aspects of video conferencing, it is also able to reorient people's faces in this way. A demonstration is available at the end of this video: https://youtu.be/eFK7Iy8enqM. But this is an expensive approach with some limitations. He et al. (2021) propose *FutureGazer*, a program which simulates eye-contact and gaze amongst participants, with readier application to existing hardware.

Another area of facial augmentation relates to dubbing of films and video, to help actors' mouths appear less out of sync with the dubbed voice. This technology dates back at least as far as 1997, with the release of Video Rewrite, "the first facial-animation system to automate all the labeling and assembly tasks required to resync existing footage to a new soundtrack" (Bregler et al. 1997: 353). his technology has since progressed significantly. In 2016, a team at the University of Erlangen-Nuremberg, the Max Planck Institute for Informatics, and Stanford University released 'Face2Face', which built on previous approaches to changing facial movements that worked 'offline' (augmenting the video more slowly than it was playing, then re-recording). Their goal was to make the process work 'online', in real time. They would train a camera on a source actor who made facial movements; the software then dynamically mapped those source movements on to the target video simultaneously. The result is the ability to change the facial movements in an existing video, in real time, by mimicking the source actor. A video demonstration is available here: https://youtu.be/ohmajJTcpNk.

Figure XX: Image: https://justusthies.github.io/posts/face2face/

Like Breglet et al. (1997) before them, Thies et al. (2016) target dubbing - "face videos can be convincingly dubbed to a foreign language" (p. 2387). And by making it work 'online', they also target real-time uses, "for instance, in video conferencing, the video feed can be adapted to match the face motion of a translator" (ibid.). This would be a human translator (as the source actor), whose facial movements Face2Face would map on to the live video of the person speaking. So there were two main limitations to Face2Face: it could only alter the mouth movements within an existing video; and it required a source actor. The next evolution of this technology would progress beyond that.

The team behind Face2Face went on to develop 'NerFACE'. This did away with the need for a source actor or indeed a video of the original person speaking. NerFACE begins by applying machine learning to recordings of a person speaking and otherwise using their face. It builds up a series of four-dimensional maps of each facial expression: everything from how that person looks when they say each vowel to the contours of their smile, their frown, and so on. With all that programmed in, NerFACE can make the recorded face say and do completely new things, achieving "photo-realistic image generation that surpasses the quality of state-of-the-art video-based reenactment methods" (Gafni et al. 2020: 1). Alongside progress with 'visemes' - face shapes associated with each sound (e.g. Peymanfard et al. 2021) - there is clear potential for anyone's face to be quickly and convincingly re-animated to say anything.

As machine translation becomes quicker and approaches real time, NerFACE could create a machine-generated

visualisation of your face speaking another language. As we discussed earlier, machine translation is not quite yet able to work in real time with your speech; but it is getting there. Progress here could also enable convincing animations of historical figures. Right now there are somewhat rudimentary re-animations using currently available methods, often for fun purposes, for example getting Winston Churchill to sing pop songs. As the technology improves, this could become more convincing, and more immersive. Imagine, for example, Da Vinci's Mona Lisa giving an art lecture, Albert Einstein teaching physics, or Charles Darwin lecturing about the science of evolution. All these are perfectly predictable extensions of these technologies currently in development. And all this will be complemented by lifelike avatars in virtual bodies, which we come back to under Virtual Reality below.

Meanwhile advances in real-time hologram technology point to a near future where augmented faces could also be beamed onto surfaces or in space around us, using the relatively low processing power of phones and other wearable devices; see for example the Tensor Holography project at MIT: http://cgh.csail.mit.edu/.

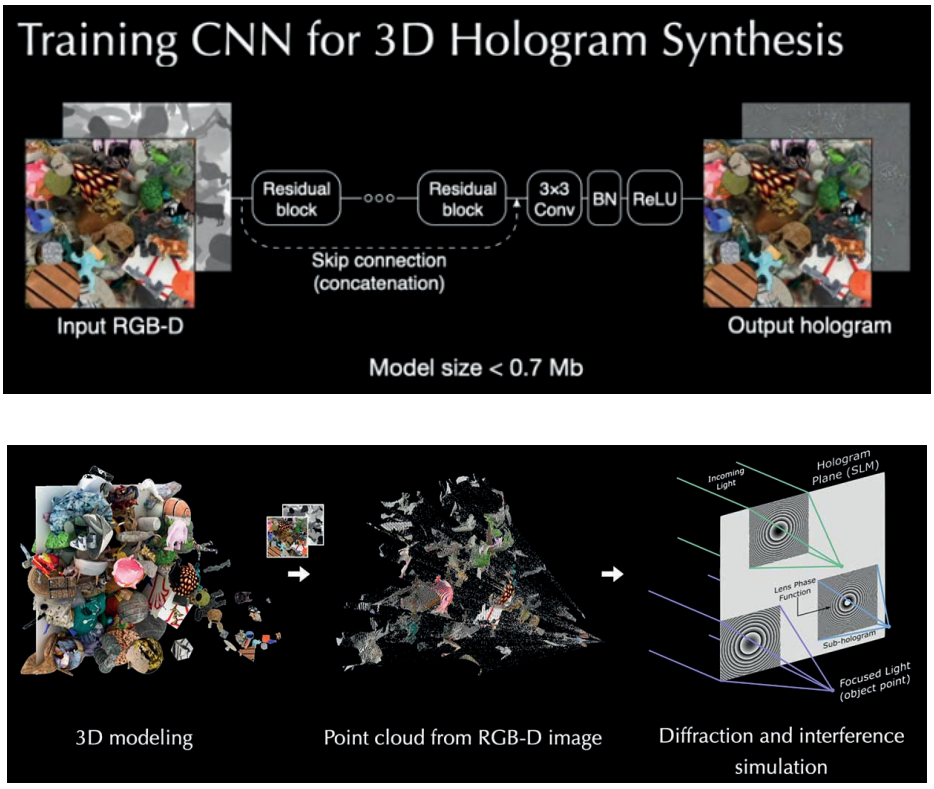


Figure XX: Images from: http://cgh.csail.mit.edu/

So, near future advances in augmented reality, combined with improved eyepieces and earpieces, suggest the following distinct possibilities within the foreseeable future:

- Talking with people in high-noise environments or over a distance and hearing their amplified and clarified voice;
- Real-time translation of people speaking different languages, fed into your earpiece as you talk to them;
- An augmented view of their face, fed into your eyepiece, so you see their mouth moving as if they were speaking your language, or if using the same language, 'seeing' their face when it is covered by clothing or they're facing away from you.

As we discussed earlier, this is unlikely to include sign language anywhere near as comprehensively in the foreseeable future. That is down to the combined problems of mixed modality in sign language (handshapes, gesture, body position, facial expression, etc.) that are much harder for machines to understand, and generally lower funding and incentives in a highly privatised and profit-led innovation space.

But the innocent uses of speech and face synthesis outlined above can and will also be used for malicious purposes, to impersonate people for profit, to spread misinformation, and to abuse people by making lifelike depictions of

them doing and saying degrading and disgusting things. This is already a reality with the rise of 'deep fakes', and indeed this has been showcased in a number of high-profile contexts, including the feature series Sassy Justice, made by the creators of South Park and featuring an augmented Donald Trump and other celebrities, doing and saying things that are out of character. The future will see this kind of augmentation available in real time on mobile equipment, with the potential for quite a mix of exciting prospects and challenging risks.

3.2 Virtual reality

The previous section covered Augmented Reality, changing things we see and hear in the real world around us. Virtual Reality (VR) is the total replacement of the real world with a simulated environment, replacing instead of augmenting what we see and hear. It typically involves wearing a bigger, opaque headset that closes off real-world stimuli and directs us entirely towards the world created by the machine.

VR is a multidisciplinary area of research and development, whose objective is to create interactive experiences/simulations mediated by computers. Stimulation of various senses allows the user to be transported to the virtual environment, which translates into a sensation of presence. In recent years, VR has gained increased interest with the release of the Oculus Rift in 2013 - later acquired by Facebook in March 2014 for $2 billion. Further new devices followed, for example HTC Vive in 2016 and HTC Focus in 2018.

Figure XX. A player using the HTC Vive (source: https://en.wikipedia.org/wiki/HTC_Vive)

Relatively affordable prices have led to high levels of adoption. AR and VR are poised to grow and mature quickly, fueled by enormous private investment; see for example Facebook Reality Labs, explicitly aiming for widespread adoption of AR and VR: https://tech.fb.com/ar-vr/. According to market analysis (Consultancy.uk 2019), VR is forecast to contribute US $138.3 billion to global GDP in 2025, rising to US $450.5 billion by 2030.

VR will be a key contributor to the human-machine era. And while AR will enable us to speak *through* technology, with VR we will also increasingly speak *to* technology. VR enables both: it can augment our voices, faces and movements; and can provide us with virtual characters to talk to.

Progress in the field may remind us of the 'uncanny valley' findings, which suggest that humanoid objects that resemble actual humans produce unsettling feelings in observers. In other words, this hypothesis suggests that there is a relationship between an object that resembles a human being and the emotional response to that object. This may be taken as a metaphor for research on previous technologies that were at first rejected but gradually embraced. This will certainly play out in the coming years.

The key advance in the coming years will be the improvement and refinement of automated conversation (chatbots), discussed in the next section below. As chatbots become more able to conduct natural, spontaneous conversations

on any topic, VR will gradually change. At present we can interact with other connected humans, or relatively simplistic chatbots. Given the likely technological advances we have discussed so far, in the near future that line will blur, at least our sense of whether we are talking to a human or a machine. That will be the key turning point for the human-machine era: talking *to* technology.

3.3 Automated conversation (chatbots)

As noted above, the human-machine era will be defined by speaking *through* technology and *to* technology. The latter will be the purview of chatbots.

Currently there are multiple platforms, thousands of startups and established businesses, and an increasing number of conferences and academic projects in the area of chatbots. Chatbots managed to become a special sort of interface with their own user experience (UX) research. According to market analysis by Fortune Business Insights (2021), the chatbot market is predicted to grow from $400 million in 2019 to $2 billion by 2027. New types of use cases have emerged for "conversational AI" in customer service: insurance, finance, retail, legal applications. Marketing as well is gradually moving from "mobile-first" (focusing on engaging users through mobile devices) to "chatbot-first". The key to this is enabling the chatbot to do things other channels cannot: going from simply answering simple questions as a text-based FAQ could do, to having a more natural conversation to determine user needs. For marketing purposes, an intelligent chatbot in a virtual assistant could suggest products and services as you speak to it, by interpreting and anticipating your needs (inVentiv Health Communications 2017). That would be the real leap from current approaches to marketing, with their comparatively clumsy reliance on your text search history, social media posts and so on.

Virtual voice assistants such as Apple Siri, Amazon Alexa, Google Assistant, Microsoft Cortana or Samsung Bixby, integrated with smartphones or smart home devices, are nowadays increasingly mainstream. Globally a total of 4.2 billion virtual voice assistants were being used in 2020. That is expected to increase to 8.4 billion by 2024 (Statista 2021). Approximately 27% of the online global population is using voice search on smartphones, with the most common voice searches being related to music (70%) and the weather forecast (64%).

Some examples of state-of the-art general-purpose chatbots include Gunrock (the winner of the 2018 Amazon Alexa Prize), Kuki (formerly Mitsuku, made by Pandorabots, the winner of the Loebner prize several years in sequence), Blenderbot (open-source conversational model provided by Facebook) and Meena (end-to-end conversational model created by Google, not open-access). The frontier for chatbots is "open-domain" use; that is, a chatbot able to talk about any topic at all without constraint.

Chatbots have also been created for more specific tasks, such as Dialogue-based Intelligent Computer-Assisted Language Learning (DICALL), medical symptom checkers, and chat-based money transfer. Various types of stakeholders are involved in production of technology in all tasks. A huge variety of tools and platforms currently support this development, and more and more new players join this domain every day.

In all use-cases, chatbots have been given one of the roles that usually a human would have in a similar type of dialogue with other humans. Take educational chatbots as an example. Due to the teacher-student role dichotomy in offline educational settings, chatbots are usually assigned the role of a teacher or a tutor. Educational chatbots have mostly been developed for second language learning, though they also exist in other domains, such as engineering and computer science.

A new use-case for chatbots emerged with the need to make ML-based systems understandable for users. Automated generation of explanations of the decisions made by complex technical systems sounds very tempting, and many researchers work on this topic (e.g. Stepin et al. 2021). Explanations may include complex interactions with embodied virtual agents as well as advanced visualizations, not only written/spoken language. In Europe, two major efforts investigate how to design and develop ML-based systems that are self-explanatory in natural language and without bias. These are the EU-funded network projects NL4XAI, https://nl4xai.eu (Alonso & Catala 2021) and NoBias, https://nobias-project.eu (Ntoutsi et al. 2020).

Various types of actors contribute to the development of chatbots. They include:

1. **Researchers** working on the underlying technology: NLU, NLG, dialogue models, user models, user experience, security, privacy, and so on.

2. **Language technology creators**. **Startups** usually offer ready-to-use packages to support chatbot development. Some of the first chatbot startups have since been acquired by bigger companies. **Big companies** offer scalable infrastructure to facilitate chatbot development and deployment (IBM, Google, Microsoft, Facebook) and integration in their own technology products (SAP).
3. **Connection providers**. Messengers offer APIs to connect the bots with their users and customers (Facebook Messenger, Slack, Kik, Viber, Telegram, etc.). But chatbots also exist in games, VR and AR applications, webchat, e-learning systems, cars, smart-home applications and many others.
4. **Policymakers**: discuss what the chatbots should and should not do.
5. **Users**: love and hate chatbots.

We discuss each of these in turn.

Researchers have worked on all aspects of chatbots since decades. While automated dialogue was rather a niche before 2015, the number of conferences dedicated to chatbots increased thereafter. Most research groups and conferences focus on dialogue systems, for example SIGDial, SemDial, CUI (focused on conversational user interfaces), CONVERSATIONS (user experience and conversational interfaces), EXTRAAMAS (explanation generation, among other topics). Major AI and NLP conferences and journals also publish about various aspects of chatbots.

Researchers create models of conversations/dialogues, find suitable algorithms to implement those models, and deploy software that uses those algorithms. Highly influential theories from the past include:

- Attention, Intentions, and the structure of discourse (Grosz & Sidner 1986).
- Speech Acts, emanating from Searle (1969) but developed for annotation in corpora (Weisser 2014).

Current approaches dominating conferences and workshops are variations of Deep Learning-based end-to-end systems trained on masses of examples (e.g. Meena and Blenderbot).

Research topics cover a variety aspects of *discourse processing* (e.g. discourse parsing, reference resolution, multilingual discourse processing), *dialogue system development* (e.g. language generation, chatbot personality and dialogue modelling), *pragmatics and semantics of dialogue*, *corpora and tools*, as well as *applications* of dialogue systems.

Language Technology Creators and Communication Channels providers are distributed across a variety of startups and big companies. Language technology companies provide business-to-business and business-to-customer conversational AI solutions. Pre-trained language models can be downloaded and reused or accessed via APIs and cloud-based services. Proprietary language understanding tools (such as Watson and LUIS) and open source alternatives (such as RASA) facilitate the development of chatbots. Services such as ChatFuel and Botsify allow creating chatbots for messengers within hours. And support tools such as BotSociety facilitate chatbot design. These tools take a great deal of programming work out of chatbot design, significantly lowering the technological bar and enabling a wide range of people to build a chatbot.

Language technology creators focus less on answering challenging research questions, more on providing a service that is scalable, available, and usable; and that satisfies a specific need of a specific customer group. For instance, NLU-as-a-Service providers only support their users in the NLU task. All remaining problems related to the production of a good chatbot are left open. Even with the support of NLU platforms, chatbot authors still need to provide their own data for re-training of usually existing language models. Especially for cloud-based services, chatbot builders need to read the privacy notice carefully: what happens with the training examples that are uploaded to the cloud? After a chatbot is developed, it needs to be connected with its users on a particular channel. Typically chatbots are deployed on instant messengers such as Facebook Messenger, Telegram, Viber, WhatsApp.

Policy makers regulate the chatbot development implicitly and explicitly. For an overview of AI policy in the EU, see https://futureoflife.org/ai-policy-european-union/. As of May 2021 the EU has 59 policy initiatives aimed at regulation of different aspects of AI - up from 51 in March 2021 (see https://oecd.ai/dashboards/countries/EuropeanUnion). Implicit regulations include for instance GDPR, which influences among other things how chatbots can record and store user input. Some new privacy regulations have negatively impacted the user experience, for example GDPR policies invoked in December 2020 removed many popular and engaging functions of online messengers, including persistent menus (previously shown to improve the user's freedom and control: Hoehn & Bongard-Blancy 2020) - disabled for all chatbots in to a Facebook page based in Europe, and/or for users located in Europe (see https://chatfuel.com/blog/posts/messenger-euprivacy-changes). The same update

disabled chatbots from sending video or audio, or displaying one-time notifications. The solution suggested by some technology provider was to create a copy of the bot with the reduced functionality and to have two versions: one bot with the full functionality, and the other "reduced European version". So the evolution of chatbots is not linear, and will be punctuated by such debates over the balance between privacy and functionality.

There are wider ethical and philosophical questions raised by speaking *to* technology in the form of chatbots; and these questions in turn highlight the challenge to many traditional areas of linguistic research. Capturing some of these question, in 2019 the French National Digital Ethics Committee (CNPEN) opened a call for opinions on ethical issues related to chatbots (https://www.ccne-ethique.fr/sites/default/files/cnpen-chatbots-call-participation.pdf). Questions included: "*Should a chatbot be given a human name?*", "*Is it desirable to build chatbots that detect human emotions?*" and "*How do you envisage chatbots influencing the evolution of language?*". Chatbots are interfaces to computer systems, and software. If we ask all these questions about chatbots, we need to ask the same questions about social networks, websites, cars and smart homes. Such is the challenge presented by new and emerging technologies, in the human-machine era, to our basic understanding of what it means to use language.

In 2019, the European Commission published a report that proposes a high-level architecture for public-service chatbots (PwC EU Services 2019). The document recommends to "use a network of chatbots, where chatbots can redirect users to other chatbots". Such formulations assign agency to chatbots that may become a source of fear of manipulation and dominance of chatbots over their users. For instance, the CNPEN questionnaire mentioned above formulates this as a very problematic question: *"... in the event of a deviation from the diet that a doctor has prescribed for a patient, the chatbot informs the doctor or even contacts the health care organization. ... do you think the chatbot's behavior is justified?"*. In fact, it is a matter of software design *and* legal regulation whether one online service (i.e. a chatbot, a website or a payment provider) can and may redirect its user to another service. Therefore, policy makers need to acquire a very sober and pragmatic view on conversational interfaces, and be aware of a potential regulation bias against chatbots.

Finally, the **users** of the chatbot technology are those whom all the stakeholders mentioned above try to please, to reach and to protect. Although earlier studies questioned the value of chatbots (Shawar & Atwell 2007), more recent research shows that chatbots are often introduced to users for inappropriate use cases and user needs are ignored, so that a negative experience with chatbots is a sort of self-fulfilling prophecy (Brandtzaeg & Følstad 2018). However, user studies show that productivity (speed, ease of use and convenience) is the main reason for using chatbots (Brandtzaeg & Følstad 2017). A focused analysis of specific cases such as retail chatbots (Kasilingam 2020) reveal that the users' attitude towards chatbots is dependent on its perceived usefulness, perceived ease of use, perceived enjoyment, perceived risk and personal innovativeness. The decision to use or not to use a bot is, however, directly influenced by trust, personal innovativeness and attitude, according to (Kasilingam 2020). These results show that an excellent user experience and a chatbot's pragmatic value need to be created with the target audience in mind: persons who are innovative on their own. This finding also makes questionable all attempts to create chatbot-based assistive technology for "elderly people" or "veterans". For sure, some of them bring the required level of innovativeness, but not all of them, and not all young people automatically adopt new technology quickly, as (De Cicco et al. 2020) showed experimentally.

With the rise of smartphones, bots like Siri, Cortana and Google Assistant reached a wider audience. However, their use was limited to a specific platform and their functions were (and still are) limited to device controls (start an app, call a person) and simple operations (book a table) only available in specific regions, mainly not in Europe.

Google developed Google Duplex (Leviathan & Matias 2018), an extension of Google Assistant that can communicate and set an appointment with a hair salon, dentist, doctor, book a flight, hotel room, restaurant table etc.

3.4 Second Language Learning and Teaching

Machine translation, electronic dictionaries and thesauri, spelling, grammar and style checkers are all helpful tools to support second-language (L2) learners. Chapelle & Sauro (2017) provide a comprehensive overview of all applications of technology for learning and teaching foreign languages. The list includes technologies for teaching and learning grammar, listening, reading, comprehension and intercultural competence.

Computer-Assisted Language Learning (CALL) at its origins was expected to facilitate learning and teaching of foreign languages by providing electronic workbooks with automated evaluation and vocabulary training.

3.4.1 Intelligent Computer-Assisted Language Learning (ICALL)

CALL technology extended with Natural Language Processing (NLP) techniques became a new research and application field called Intelligent Computer-Assisted Language Learning (ICALL). Language technology has been integrated into CALL applications for the purposes of automated exercise generation (Ai et al. 2015), complex error analysis and automated feedback generation (Amaral 2011).

Petersen (2010) differentiates between Communicative ICALL and Non-Communicative ICALL. He sees Communicative ICALL as an extension of the human-computer interaction field. His understanding of Communicative ICALL is that "Communicative ICALL employs methods and techniques similar to those used in HCI research, but focuses on interaction in an L2 context" (Petersen 2010: 25). We look at applications in both subfields.

Frequently cited real-life ICALL applications are E-Tutor for German learners (Heift, 2002, 2003), Robo-Sensei for Japanese learners (Nagata 2009) and TAGARELLA for learning Portuguese (Amaral et al. 2011). These systems have conceptually similar structures.. Main technical components include an expert model, a student model and an activity model. Language technologies are heavily employed to analyse learner errors and provide corrective feedback. Automated linguistic analysis of learner language includes the analysis of the form (tokenization, spell-check, syntactic parsing, disambiguation and lexical look-up) and the analysis of the meaning (whether the learner answer makes sense, e.g. expected words appear in the input, the answer is correct etc.).

Typical learner errors are usually part of the instruction model in ICALL systems. Corpora of learner language are used to model typical L2 errors for different L1 speakers, for example FALKO (Reznicek et al. 2012, 2013), WHiG (Krummes & Ensslin, 2014) and EAGLE (Boyd 2010). The repository Learner Corpora around the World contains many other learner corpora. The annotation of learner corpora is mainly focused on annotation of learner errors; however, annotation of linguistic categories in learner corpora is also of interest. Error annotation of a corpus assumes a non-ambiguous description of the de- viations from the norm, and therefore, the norm itself. The creation of such a description may even be problematic for errors in spelling, morphology and syntax (Dickinson & Ragheb 2015). In addition, different annotators' interpretations lead to a huge variation in annotation of errors in semantics, pragmatics, textual argumentation (Reznicek et al. 2013) and usage (Tetreault & Chodorow, 2008). Multiple annotation schemes and error taxonomies have been proposed for learner corpora, for instance (Díaz-Negrillo and Domínguez, 2006; Reznicek et al. 2012).

3.4.2 Communicative ICALL

Chatbots (dialogue systems) are already used in language learning environments, both for practising written and spoken dialogues in everyday scenarios, e.g., Duolingo chatbot characters that give contextual and unique responses to users instead of the same response for all similar inquiries. The idea of employing chatbots as language learning tools is rather old (see e.g. Fryer & Carpenter 2006). Many benefits of this have been reported – amusement and engagement in learning, reduced anxiety compared to talking to a human, availability at any moment, possibility of repetition (the bot doesn't get bored) and multimodality (practising reading-writing and listening-speaking skills simultaneously). Communicative ICALL is also called "Dialogue-based ICALL".

Education, industry and language learners benefit from various deployed applications, for instance Babbel (Heine et al., 2007) Alelo (Sagae et al. 2011) and Duolingo (von Ahn and Hacker, 2012). A number of mobile applications in the AppStore and GooglePlay Store target conversation training and traditional task-based language instruction. Conversational agents, chatbots and dialogue systems for foreign language training have been developed as stand-alone conversation partners (Jia 2009; Timpe-Laughlin et al. 2017) and as part of intelligent tutoring systems (ITS) (Petersen 2010), serious games (Sagae et al. 2011; Wik et al., 2007~~, 2009~~; Amoia et al. 2012) and micro-worlds (DeSmedt 1995). Recent overviews in Communicative ICALL are provided in (Bibauw et al. (2019) and Hoehn (2019: Ch.2).

Putting a chatbot or a robot in the role of a teacher may lead to a power conflict: in a standard classroom role distribution, a teacher has more power than a student. This has been well-researched for instance in conversation-analytic studies of L2-classroom interaction (see for instance Kasper & Wagner 2001, and Markee 2000). In interaction with machines, humans argue that machines never should become bosses of humans, machines never should be allowed to control human beings (Asimov 1941). Therefore, it is not surprising that at present, language learners are likely to prefer a human tutor to read and assess their text. Although arguments exist that

more intelligent software can easily change this perspective (Ai 2017)), the intelligence of the machines will not solve the problem of power. Nevertheless, making tutoring chatbots more "human-like" while assigning a new role to them (not similar to a teacher) may help to resolve authority problems. Two examples of such roles have been explored in (Hoehn 2019) and in (Vijayakumar et al. 2018). The former explores using chatbots that simulate native speaker peers for practicing L2 conversation (equal-power relationship). The latter describes how chatbots can be a teacher's helpers and support beginner-learners with interactive practicing of software engineering concepts and personalised formative corrective feedback (a chatbot does not replace the teacher).

3.4.3 VR-based Language learning

VR is already widely used for language learning (see Parmaxi 2020; Wang et al. 2020b), by providing learners with scenarios that mimic real-world situations and conversations. This immersive language learning method has many advantages, being one of the most valuable possibilities of practising in a context that lets users engage with the language and culture, and indeed with imaginary situations and cultures that provoke interest and engagement. The immersive multiplayer game Second Life is used by many universities and language learning organisations to facilitate learning. There is an informative Wikipedia page covering these: https://en.wikipedia.org/wiki/Virtual_world_language_learning

Educators have the recurring challenge of maintaining the students engaged in the learning process, and the teaching of a foreign language is no exception. In this sense, VR technology can play a key-role to introduce new techniques and stimulate learners' motivation. VR offers a unique capability of transporting users to a virtual space and immersing them in there to such a level that they feel that they are really there as if they were physically in a real location. Since 2004, the number of works has been growing consistently (Lin & Lan 2015; Solak & Erdem 2015; Parmaxi 2020). Together these show a consensus about a positive interaction between the usage of VR for learning and learning outcomes. Some examples of previous work that established a link between the feeling of being present in the VE (i.e. the sense of presence) and the learning outcomes, where a higher sense of presence contributes positively to learning (North & North 2018; Makransky et al. 2017). An explanation for this phenomena is that due to users feeling more present and engaged with the VE they devote their attention to the VE and this focus allows them to develop the concentration required for absorbing new knowledge.

In particular, the VR environment enables us to incorporate kinesthetic learning in language education. Research shows that people remember more, up to 90%, by doing things (Dale 1946). This has been confirmed by studies in the field of neuroscience. Repetto (2014) maintains that a virtual motion (a motion performed in the virtual world with a body part that is actually steel) associated with action words can enhance verbal memory if the environment is seen as real-life. Furthermore, Vázquez et al. (2018) reveal in their research study that virtual kinesthetic learners exhibited significantly higher retention rates after a week of exposure than all other conditions and higher performance than non-kinesthetic virtual reality learners. Moreover, Tseng et al. (2020) show that VR mediation positively affected students' achievement results provided that students worked autonomously in the individual and paired work conditions. Legault et al. (2019) expand that the VR environment is particularly suited for less successful learners who in their study on non-native language vocabulary profited more than successful learners of second language. Generally, it seems that students improve their proficiency level by using VR technologies (Hassani et al. 2016).

Focusing on immersive VR setups, previous work has already addressed the adoption of such languages from the perspectives of both teachers and students. Both of them agree that VR can play a valuable role in the learning process (Hansen & Petersen 2012). For teachers, VR is seen as a new and valuable tool to engage students by converting the theoretical concepts into these new practices that can facilitate the student's learning (Wang et al. 2015; Andresen & van den Brink 2013, Peixoto et al. 2019).From the students point-of-view, there are two approaches: the use of these technologies as complementary to the ordinary curriculum or as an application that promotes autonomous learning, often by means of gamification. When used as a complementary tool to the regular curriculum, students have revealed not only better learning outcomes but also more satisfaction and motivation to learn the new language (Ijaz et al. 2017). As for the usage of VR for autonomous learning reinforced with gamification strategies, surprisingly it has been shown to have superior learning progress when compared to other methods (Guerra et al. 2018; Barreira et al. 2012). This is explained by the fact that gamification learning is able to create an additional engagement as it appeals to the competitive nature of the learners to progress in the game and to unlock bonus features or achievements (Buckley & Doyle 2016). In addition, Chen & Hsu (2020)

suggest that the interaction feature of the VR application and the challenges of game-based design enable students to enter the state of flow easily and enhance their motivation to learn.

VirtualSpeech-VR Courses – this app is not aimed primarily on non-native language learning, but it mainly develops business skills. As Symonenko et al. (2020) claim it can be successfully exploited in business English language courses. In addition, its function of speech analysis enables students to get feedback on their speeches, record all the speeches and have the progress results.

VR Learn English App – this app is aimed at learning English as a second language. In fact, this app enables a learner to walk inside the house with its different rooms and objects. The learner walks around various objects and by pointing them with his/her VR headset, the app provides its corresponding name and shows it on the learner's screen. The learner listens to the pronunciation of the word, repeats it and memorizes it.

Mondly VR: Learn Languages in VR – this app focuses on non-native language learning. Moreover, it operates in 33 different languages. It has two sections: a vocabulary section and a conversation section. A learner studies new words and phrases in context (e.g. in a restaurant or at the post office), practices other language skills, i.e. listening and reading, as well as s/he is provided feedback on his/her pronunciation.

InmerseMe – an online language learning platform that contains the teaching of several languages which offers several configurations based on VR. The use of this technology allows a most authentic representation of the region where the desired language is spoken. After selecting a configuration and a class, students interact with pre-recorded "teachers". Beginners can participate in dictation exercises in which they repeat the words spoken by the virtual teachers. The words spoken by the students are automatically recorded and transcribed by the application. The application provides materials for learning nine languages. Briefly, the application creates a contextualized environment in which students can improve their skills in the desired language.

Panolingo – this application offers users the opportunity to learn a foreign language through the gamification approach, where they can earn points and bonuses. You can share your scores with friends on the app to stimulate a competitive mindset and engage users in the learning process. The immersive nature of this interactive platform offers guidance to users to complete different challenges, providing more effective learning experiences as it exposes the users directly to scenarios where the foreign language is the basis of the whole scene.

The descriptions of the apps mentioned above indicate that the only skill, which is not practiced, is the skill of wiring. Otherwise, the content of the app attempts to teach the target language in context through meaningful activities with a special focus on the development of vocabulary. However, not much attention is paid to collaborative learning, which appears to be crucial for learning a non-native language because it facilitates understanding, develops relationships or stimulates critical thinking.

Foreign/Second Language learning through VR presents interesting didactic challenges. Its educational potential still needs to be addressed and researched thoroughly as it is yet to be presented as a better didactic tool/solution compared to traditional classroom strategies and approaches (Hansen and Petersen 2012), despite VR´s much-studied benefits and positive learning outcomes (Lin ~~and~~ Lan 2015, Parmaxi 2020, Guerra et al. 2018 and Barreira et al. 2012, Solak & Erdem 2015).

Despite the growing interest and number of works made regarding the use of VR technologies in learning, there is still much work to be carried in the field to identify how the VR technology can be exploited thoroughly. Despite the majority of the literature points towards a positive impact of VR technologies, other studies suggest that in the long term, conventional approaches are better than technologically-based approaches (Sandusky 2015). More extensive and longitudinal studies are needed to understand the long-term impact of the usage of VR technologies in the teaching/learning processes.

Likely future developments

Immersivity can be enhanced further by producing multisensory inputs (sound, sight, touch), since this enhances learning. In a systematic review of the literature in this area, Parmaxi (2020) forecasts various likely avenues for development in VR language learning, including: "Real-life task design … For example, the development of real-life tasks within a virtual world would allow for authentic language production in situations in which language learners would need to use the language, rather than performing mechanic, fill-in-the-gap activities" (Parmaxi 2020: 7).

VR language learning can be used for human teachers or conversation partners to interact with students in a virtual space. But in due course, as virtual avatars improve and become more realistically lifelike - incorporating advances

in chatbots discussed elsewhere in this report - so too this new technology will allow language learners to speak to virtual teachers and conversation partners. This has notable benefits. Learners may feel more comfortable practising their accent and pronunciation without feeling embarrassed, and could receive instant feedback at low cost, or in exchange for their usage data. But it may come with risks and challenges: perhaps some learners would struggle to develop the confidence to go on to seek with actual humans; and there will be exclusion based on accessibility and digital literacy. Again, foresight is needed, and research must be built into the design of new systems.

Further key questions still need to be addressed. Is second language learning better taught by means of VR? Should traditional methodologies and approaches be completely replaced? Who should be the target learners? Is it an universal method? Surely, these are some questions whose answers are much needed to come forward with significant results and conclusions.

3.4.4 AR-based Language Learning

While everything in VR settings is "computer generated", AR enables the learner to interact with the physical/real world superimposed by virtual objects and information. AR is very good at overlaying the world with words: flashing up road names, opening times, factual information, and in the near future more complex information like conversation prompts and augmented conversation (as we discussed earlier). When combined with advances in machine translation, this is absolutely ripe for language learning (Zhang et.al. 2020). Through the use of mobiles, tablets and wearable technologies like smart glasses and contact lenses, AR will enable innovative learning scenarios in which language learners can use their knowledge and skills in the real world. Learning by doing.

AR has a relatively recent history of research and application in second language learning. In the relevant literature, recurring themes regarding the potential benefits of AR applications are: a) learning through real-time interaction, b) experiential learning, c) better learner engagement, d) increased motivation, e) effective memorization and better retention of the content (Khoshnevisan & Le, 2019; Parmaxi & Demetriou, 2020).

Communication and interaction - vital in language learning - are effectively supported in AR-enriched environments. Through the use of QR codes, markers and sensors such as GPS, gyroscopes, and accelerometers, any classroom can be turned into a smart environment in which additional information is added to physical surroundings. These smart environments can be exploited in the form of guided digital tours or place-based digital games in which learners can interact with the objects and people synchronously, while fulfilling any language-related tasks.

If well-designed, AR-enriched environments will enable learners to take an active role speaking, reading, listening and writing in the target language. The project Mentira (Holden & Sykes, 2011) is an early example of such an environment in which a place-based game, a digital tour and classroom activities were combined to teach Spanish pragmatics. While framing requests or refusals, students were able to interact with game characters as well as actual Spanish speaking inhabitants during the tours. These innovative interactive environments also boost learner motivation and engagement.

Collaboration, novelty of the application, feeling of presence, and enjoyable tasks through playful elements are among the reported factors attributed to AR enhancing learner motivation (Liu & Tsai, 2013; Chen & Chan, 2019; Cózar-Gutiérrez ~~and~~ Sáez-López 2016). Experiential learning - the "process whereby knowledge is created by the transformation of the experience" (Kolb, 1984) - is another theme frequently underlined in AR-based language learning. Experiential learning emphasizes that learning happens when learners participate in meaningful encounters through concrete experience and reflect upon their learning process.

AR-based instruction has potential to turn the educational content into a concrete experience for the learners. Place-based AR environments, for example, guide the users at certain locations and help them to carry out certain tasks through semiotic resources and prompts. Within the language learning contexts, these tasks could be in the form of maintaining a dialogue at an airport or a library or asking for directions on a street. As the learners' attention is oriented to relevant features of the setting, their participation into the context is embodied, which makes their experience more concrete than in-class practicing of these tasks. ~~This "embodied participation into language learning process" (Wei, 2018 p.18)~~ also leads to successful uptake and better retention of the content .

Based on the possible benefits above, there is an increasing interest in AR tools and applications integrated into language education, which yields more and more attempts in developing new tools or adapting existing ones into educational settings. Below is the brief presentation of the AR software frequently used for language teaching purposes.

ARIS (Augmented Reality and Interactive Storytelling), developed at the University of Wisconsin. As a free open-source editor, ARIS could be used by any language teacher without technical knowledge to create simple apps like tours, scavenger hunts. More complex instructional tasks, on the other hand, require HTML and JavaScript knowledge. Action triggers could be either GPS coordinates (i.e. location) or a QR code, which could be exploited to start a conversation, seeing a plaque or visiting a website to guide learners in their language practice (Godwin-Jones, 2016). Using ARIS, Center for Applied Second Language Studies at the Oregon University has recently released a number of language teaching games and free to use AR materials targeting different proficiency levels, which means language teachers could readily integrate AR into their teaching practices.

TaleBlazer, developed at MIT. Its visual blocks-based scripting language enables the users to develop interactive games avoiding syntax errors. GP coordinates could be used as triggers which could initiate certain tasks targeting speaking or writing or practicing vocabulary in the target language. Imparapp developed through this software at Coventry University to teach beginner level Italian is a good example of exploiting TaleBlazer for language teaching purposes (Cervi-Wilson & Brick, 2018).

HP Reveal Aurasma. A free to use tool, HP Reveal Aurasma has both mobile and web interfaces. While the mobile version is generally used to view AR materials and offers limited opportunity to create AR materials, the web interface (i.e. Aurasma Studio) provides the users with a wide range of tools including content management, statistics and share options to develop any AR enriched environment. A myriad of ready to use AR material that any teacher would take to their classrooms make it one of the frequently preferred tools by the teachers. Its use for language teaching purposes range from creating word walls to teach vocabulary, designing campus trips to interactive newspaper articles to improve reading skill in the target language. ~~For inspiring ideas regarding the Aurasma for language teaching, please see Yno, 2010; Antonopoulos, 2016; Driver, 2016; Richardson, 2016).~~

Unity (ARToolkit), released by the University of Washington. As an open-source AR tool library with an active community of developers, Unity ARToolkit's two distinguishing aspects are efficient viewpoint tracking and virtual object interaction; these make it popular among AR developers. However, as for the teachers with less programming knowledge, it might not be so favourable.

Vuforia. As a freely available development kit (on the condition of inserting vuforia watermark), Vuforia allows the users to develop a single native app for both Android and IOS along with providing them with a number of functions including Text Recognition, Cloud Recognition, Video Playback, Frame Markers etc. It smoothly works in connection with Unity 3D and its developer community constantly offers updates, which contributes to its popularity. Although it is widely used in different fields of education, the field of language teaching has yet to embrace vuforia in instructional activities. Interested readers could examine Lin et.al. (2021), Rahaman et.al. (2021), Dalim et.al. (2020) to catch a glimpse of ideas about ways of exploiting Vuforia for language teaching purposes.

Looking ahead, the effects of an increased use of technologies (as integrated with our senses rather than simply confined to mobile or external devices) are very hard to predict regarding the cognitive treatment of the information. Research has shown that reading modalities, for instance, can impact metacognitive regulation (Ackerman & Goldsmith 2011; Baron 2013; Carrier et al. 2015) and higher order cognitive treatment (Hayles 2012; Greenfield 2009; Norman & Furnes 2016; Singer & Alexander 2017). The addition of numerous additional layers of multimodal information will have to be studied carefully to make sure that optimal cognitive treatment is possible. In case it is not, efforts will be needed to tailor-make the input received to make it manageable for the human brain.

3.5 Law and Order

This subsection discusses applications of Legal Informatics to the treatment and processing of language, including the use of technology in forensic contexts to assist: (a) legal practitioners (e.g. judges, lawyers, regulators, police officers, etc.) in their daily activities, so as to help the courts make informed, fair and just decisions; and (b) the society in general, in matters that are of common interest (e.g. detection and analysis of academic plagiarism, deception detection, etc.).

3.5.1 Automated legal reasoning

One of the first applications of artificial intelligence was to emulate the reasoning of mathematicians (McCorduck

2004). The ability of the *Logic Theorist* (Allen and Simon 1956) to automatically prove mathematical theorems is based on its use of the same logic and same reasoning procedures as those applied by mathematicians. The similarity of machine reasoning to human reasoning has various advantages:

1. The reasoning process can be traced and understood.
2. The same logic and reasoning procedures can be applied to different domain problems and are not designed for solving specific problems.
3. Such tools can support users in their own, human reasoning process

Laws are written with an intended logic behind them. Their interpretation is intended to be realized using specific reasoning procedures. Nevertheless, the logic behind laws is far from trivial, and therefore hard to program. Take for example the rich discussion about which logic is most suited to capture contrary-to-duty obligations (Prakken and Sergot 1996) and the many paradoxes demonstrating the weaknesses of various suggested logics (Carmo and Jones 2002).

Similarly, the legal reasoning process itself is very complex. Legal sentences have various interpretations, which are based on various factors, such as context and locality; these need to handle contradicting interpretations from different legal codes and types. It is not a surprise, therefore, that automated legal reasoning had only a minimal impact so far on the legal profession. Nevertheless, progress has been made on applying automated reasoning to the legal domain.

Various attempts have been made to use logics to capture different legislations. An example is the DAPRECO knowledge base (Robaldo et al. 2020), which has used an input-output logic (Makinson and van der Torre 2003) to capture the whole European general data protection regulation (GDPR).

There were also successful implementations of legal reasoning procedures. One example is the Regorous system [Governatori 2015] which is capable of checking regulatory compliance problems written in defeasible deontic logic (Governatori et al. 2013).

In the remainder of this subsection, we will describe two applications of these and similar systems to law.

Consistent legal drafting - In computer programming, a program written in a formal programming language can be executed by a computer. The compilation/interpretation process which is performed also checks the program for consistency issues and errors. The legal language has some similarities to programming languages - both depend on a relatively small vocabulary and precise semantics. In contrast to computer programs, legislations cannot be checked by a compiler for consistency issues and errors. There might be many reasons for such errors, ranging from syntactical errors in the legislation to different contradicting legislations and authorities which apply at the same time. In order to validate and check legislations for such errors, two processes are needed. The ability to translate a legislation to the computer program, for example by annotations [Libal and Steen 2019] or even manually. Once a legislation is translated, other programs can check the legislation for errors and consistency issues [Libal and Novotna 2020].

Regulatory compliance checking - Similarly, the ability to translate a legislation to a computer program has other benefits. An important application is the ability to check various documents for regulatory compliance. Regulatory compliance checking in many domains, such as the financial one, is a very complicated process. Such a process should result in a yes/no answer. Nevertheless, the process normally results in only a decreased likelihood of compliance violation. A computer program which can instantly check all possible violations and interpretations, for example when checking for GDPR compliance [Libal 2020], can greatly improve this process.

3.5.2 Computational forensic linguistics

In the human-machine era, computer forensics has attracted the attention of both forensic scientists and the general public. In recent years, the general public gained the perception that not only is scientific evidence the only evidence to take into account in forensic cases, but also the belief that such evidence can be obtained quickly and easily by using computational tools - a phenomenon that has been dubbed 'the CSI effect'. However, the field also gained a place of its own among the growing range of forensic sciences, given the range of investigative possibilities that it has to offer.

Computer forensic can, however, refer to two different applications: the forensic analysis of hardware (and installed software) in cases where computers or other computerized technologies have been used in illegal practices (e.g. criminal communications, money laundering, ...); or to the use of computational tools, methods and techniques for purposes of investigating forensic cases. In this case, computer forensics can assist DNA tests, fingerprint analysis, or ballistics, among others. Machines are indispensable to help forensic scientists process and interpret forensic data timely and efficiently. The use of computational tools, methods and techniques to analyse linguistic data, e.g. in cases of disputed authorship of anonymous texts, is dubbed **computational forensic linguistics**.

The field of computational forensic linguistics has traditionally been headed by computer scientists (e.g. Woolls 2021; Sousa-Silva 2018), while linguists have played a secondary role. They also tend to look at the same problem from different angles: computer scientists have been more interested in handling high volumes, train samples of known origin, and then test the methods developed based on precision, recall and F1 rates; for forensic linguists, computational tools are crucial to assist the analysis of the data, but determining precision, recall and F1 does not suffice to make informed decisions. For instance, for computer scientists an F1 score of, say, 85% can be excellent; for linguists working in forensic cases, this would mean that there is a 15% chance that the computational analysis is wrong. This, in forensic cases, could lead to an innocent being imprisoned, or a criminal being released. Hence, the linguists' preference for linguistic analysis, to the detriment of the computational approach. Notwithstanding, in recent years forensic linguists have acknowledged the potential of computational analyses in forensic settings (see e.g. Sousa-Silva et al. 2011; Grieve et al. 2018), so the provision of forensic linguistics expertise is increasingly seen as indissociable from a lower or higher degree of computational analysis.

The following are some of the most common applications of computational forensic linguistics:

- **Forensic authorship analysis**, often referred to also as authorship attribution, consists of establishing the most likely author of a text from a pool of possible suspects. This analysis is applied to texts of questioned authorship, such as anonymous offensive or threatening messages, defamation, libel, suicide notes of questioned authorship or fabricated documents (e.g. wills), among others. Authorship analysis typically involves identifying an author's writing style and establishing the distinction between this style and that of other authors. A high-profile case involving authorship analysis is that of Juola (2015), who concluded that Robert Galbraith, the author of the novel 'The Cuckoo's Calling', was indeed J.K. Rowling, but forensic authorship analysis has also been used in criminal contexts, e.g. the 'unabomber' case.
- **Authorship profiling** consists of establishing the linguistic persona of the author of a suspect text, and is crucial in cases where an anonymous text containing criminal content is disseminated, but no suspects exist. It allows linguists to find elements in the text that provide hints to the age range, sex/gender, socioeconomic status of the author, geographical origin, level of education or even whether the author is a native or a non-native speaker of the language. When successful, authorship profiling allows the investigator to narrow down the pool of possible suspects. Recent approaches to authorship profiling include profiling of hate speech spreaders on Twitter (https://pan.webis.de/clef21/pan21-web/author-profiling.html).
- **Plagiarism** is a problem of authorship - or, more precisely, of its violation. Although plagiarism detection and analysis is approached differently from authorship attribution, in some cases authorship attribution methods can also be helpful. This is the case, in particular, when the reader intuits that the text does not belong to the purported author, but is unable to find the true originals. An intrinsic analysis can be used, in this case, to find style inconsistencies in the text that are indicative of someone else's authorship. The most frequent cases of plagiarism, however, can be detected externally, i.e. by comparing the suspect, plagiarising text against other sources; if those sources are known, a side-by-side comparison can be made, otherwise a search is required, e.g. using a common search engine or one of the so-called 'plagiarism detection software' packages. Technology plays a crucial role in plagiarism detection; as was argued by Coulthard & Johnson (2007), the technology that helps plagiarists plagiarise also helps catching them, and an excellent example of the potential of technology is 'translingual plagiarism' detection (Sousa-Silva 2013, 2021): this is where a plagiarist lifts the text from a source in another language, machine-translates the text into their own language and passes it off as their own. In this case, machine translation can be used to revert the plagiarist's procedure and identify the original source.

- **Cybercrime** has become extremely sophisticated, to the extent that cybercriminals easily adopt obfuscation techniques that prevent their positive identification. However, cybercriminal activities, including stalking, cyberbullying and online trespassing usually resort to language for communication. A forensic authorship analysis and profiling of the cybercriminal communications can assist the positive identification of the cybercriminals.
- **Fake news** has increasingly been a topic of concern, and has gained relevance especially after the election of Donald Trump, in the USA, in 2016 and Bolsonaro, in Brazil, in 2018. The phenomenon has been approached computationally from a fact-checking perspective; however, not all fake news are factually deceiving (Sousa-Silva 2019), so they pass the fact-checking text. Given the ubiquitous nature of misinformation, a computational forensic linguistic analysis is crucial to assist the detection and analysis of fake news. This is an area where the Human-Machine relationship is particularly effective, since together they are able to identify patterns that are typical of pieces of misinformation.

3.5.3 Legal chatbots

Chatbots are nowadays present in a number of industries, including the legal one, to expedite tasks, optimize processes and ease pressure on human workers like lawyers, who can hence dedicate their time and cognitive resources to more complex matters. Question-and-answer dialogs offer the advantage of iteratively collecting the data necessary to automatically generate specific legal documents (e.g., contracts) and to offer legal advice 24/7 based on fill-in-the-blanks templating mechanisms (Dale 2019).

Filing paperwork for the general public is one of the concrete applications of legal chatbots: for instance DoNotPay (donotpay.com), widely termed the "first robot lawyer", is an interactive tool meant to help members of the US population to e.g., appeal parking tickets, fight credit card fees and ask for compensation, among the many use cases. SoloSuit (solosuit.com) efficiently helps people that are sued for a debt to respond to the lawsuit. Australia-based 'AILIRIA' (Artificially Intelligent Legal Information Research Assistant, (ailira.com/build-a-legal-chatbot) is a chatbot implemented directly on Facebook Messenger that offers the possibility to build one's own legal chatbot to advise clients on a variety of matters and promises to have a "focused and deep understanding of the law". PriBot (pribot.org/bot) is able to analyse privacy policies published by organizations and provide direct answers concerning their data practices (e.g., does this service share my data with third parties?). In such a scenario, chatbots spare individuals (end-users and supervisory authorities alike) the effort of finding information in an off-putting, lengthy and complex legal document. Chatbots are also employed to automatically check for the state of compliance of an organization with applicable regulations and even suggest a course of action based on the answers (e.g., GDPR-chatbot.com).

In all these cases, however, the tradeoff is between ease of use and low or no cost on the one hand, and reliability and formal assurance on the other. Naturally, bots that interpret legal terms and conditions also themselves have terms and conditions. Bespoke professional legal advice from a human may still have value in the human-machine era.

3.6 Health and Care

Application of voice in healthcare is a rapidly growing area of research, especially for diseases that are accompanied with voice disorders. There is ongoing work into early diagnosis of various mental and neuro-degenerative diseases, detecting subtle changes in speech that could indicate neurological decline - changes so subtle that friends and relatives typically miss or misinterpret them but which machine learning can accurately identify based on large corpora of known sufferers.

Recent attempts at the University of Cambridge (Brown et al., 2020) and MIT (Laguarta et al., 2020) were made to use speech and respiratory sounds such as coughs and breathing to diagnose Covid-19.

Vocal biomarkers can be used as audio signatures associated with a particular clinical outcome, and can be utilised to diagnose disease, and subsequently monitor its progression. Features that affect articulation, respiration and prosody are used to track the progress of multiple sclerosis (Noffs et al. 2018). The automatic detection of

Parkinson's disease from speech and voice has already entered a mature stage after more than a decade of active research, and could eventually supplement the neurological and neuropsychological manual examination to diagnose (Shahid & Singh 2020).

Linguistic analysis may characterize cognitive impairments or Alzheimer's disease, which are manifested by decreased lexical diversity and grammatical complexity, loss of semantic skills, word finding difficulties and frequent use of filler sounds (Nguyen et al. 2020; Robin et al. 2020). Because linguistic disabilities have different reasons (Antaki & Wilkinson 2012) and different effects (traumata, autism, dementia and others), different types of technology are needed to support speakers suffering from those disorders, and speakers who communicate with people with interactional disabilities.

With the development of audio devices that offer improved interoperability with Internet of Things (IoT) services, such as Amazon Echo or Google Home, integration of voice technologies into Ambient Assisted Living (AAL) systems started to get a lot of traction in the recent years. AAL systems aim at supporting and improving life quality of elderly and impaired people by monitoring their physiological status, detecting/preventing falls or integration with home automation and smart home solutions.

Vocal User Interfaces (VUI) are commonly part of such solutions since they can enable voice interaction with people with reduced mobility or in emergency situations. However, their use imposes several challenges including distant speech recognition and requirement to memorize voice commands and/or pronounce them in a particular way. The voice also degrades with aging, or can be affected by other diseases, which may affect speech recognition and performance of VUIs. There are attempts to create systems that are adapted to (possibly disordered) speaker voice and allow the users to express the voice commands freely, not restricted to particular words or grammatical constructs (e.g. Despotovic et al. 2018).

This would be entirely feasible as an embedded feature of any smartphone, virtual assistant, or other listening device. Viable widely used applications are therefore to be expected in near future. But these will of course bring major dilemmas in terms of privacy and security, as well as potential uses and abuses for providers of private health insurance and others who may profit from this health data.

Aside from diagnosing mental illnesses and cognitive decline, there are numerous AI robots and chatbots made for neurodivergent conditions, for example bots designed to help autistic children to develop and refine their social and interpersonal skills (e.g. Jain et al. 2020).

3.7 Sign-language Applications

A plethora of applications have been developed to offer a range of capabilities, including sign language generation through the combination of sign language formalization and avatar animation. In their survey, Bragg et al. (2019) articulate the challenges of sign language recognition, generation, and translation as a multidisciplinary endeavour.

As we have noted so far in this report, sign language is in fact much more than just shapes made with the hands; it also relies heavily on facial expression, body posture, gaze, and other factors including what signers know about each other. But, as we have also noted so far, progress in machine recognition of sign is currently limited entirely to detecting handshapes. Progress with handshapes alone will not help signers anywhere near as much as the kinds of technologies we have reviewed for hearing people. That is the fundamental inequality to remember in any discussion of sign language technology.

For machines to understand handshapes, there have been broadly two different approaches: "contact-based systems, such as sensor gloves; or vision-based systems, using only cameras" (Herazo 2020). UCLA bioengineers have designed a glove-like device that can translate signs into English speech rapidly using a smartphone app (Zhou et al. 2020). These wearable tracking sensors have been criticised as awkward, uncomfortable, and generally based on a premise that deaf people should deal with such annoying discomfort for the benefit of 'assistance'. No hearing person would tolerate this, much less think it was a positive addition to their life (Erard 2017).

Figure XX: Image: Erard (2017)

The second broad approach mentioned above, camera-based systems, has seen some slow incremental progress recently. For example, Herazo (2020a) shows some accurate recognition capabilities, though overall the results are "discouraging" (see also Herazo 2020b). Google's MediaPipe and SignAll (https://signall.us/sdk) embed visual detection of sign into smartphone cameras or webcams. However, as they note somewhat quietly in a blog post (https://developers.googleblog.com/2021/04/signall-sdk-sign-language-interface-using-mediapipe-now-available.html), this is still limited to handshapes, excluding the various other multimodal layers also essential to meaning in sign:

> "While sign languages' complexity goes far beyond handshapes (facial features, body, grammar, etc.), it is true that the accurate tracking of the hands has been a huge obstacle in the first layer of processing – computer vision. MediaPipe unlocked the possibility to offer SignAll's solutions not only glove free, but also by using a single camera."

It is positive to see a less physically intrusive solution; but as they briefly acknowledge, this shares precisely the same limitations as its predecessors, and does nothing to advance a solution.

Likely future improvements

Research in the area is ongoing. One example is the research conducted by the LISN (formerly LIMSI), sign language NLP group (https://www.limsi.fr/en/research/iles/topics/lsf), whose Rosetta 2 project aims at developing an automatic generator of multilingual subtitles for television programs and Internet video content for the deaf and hard of hearing. This system is based on artificial intelligence and on the development of an automatic French Sign Language (LSF) representation system in the form of animation based on a 3D sign avatar. Additionally, the group has focused on the study of "deverbalization", a key process in translation which consists of interpreting the meaning of a message independently of the sum of the meaning of the individual words used, so as to develop a software tool for assisting sign language translators. Sign language MT is still significantly underdeveloped, when compared to machine translation of verbal language, and likely to remain behind for the foreseeable future.

As noted above, more (and bigger) sign language corpora are needed. Sign-hub, the project discussed above, has made some advances in creating blue-prints for sign translation, which, as noted, allows some progress, but with tight limitations. Significant further advances in automated sign translation will require corpus-based basic research comparing different sign languages, so as to be able to show variation in language use. So far, this type of comparative research has yet to begin in earnest.

With all the above in mind, for the foreseeable future, automated sign is likely to be limited to specific use case interactions, and one-way translation of basic messages, for example a program allowing for the translation of train traffic announcements from French into French Sign Language, displayed on monitors in the railway station (Ségouat 2010). The development of a universal, robust and cost-effective sign language machine translation system is currently unrealistic (Jantunen et al. 2021); even more so if we consider all the potential use cases involving sign language translation. And if gloves or even cameras improve, this will still remain notably more awkward than the sleek glasses and earbuds that lie ahead for hearing people.

Further projects and organisations to watch for future progress in this area include:

- Signing Avatars & Immersive Learning (SAIL) at Gallaudet University, USA, funded by the US National Science Foundation, nsf.gov/awardsearch/showAward?AWD_ID=1839379
- The American Sign Language Avatar Project at DePaul University, http://asl.cs.depaul.edu
- Simax, the sign language avatar project, by Signtime GmbH, a company in Austria, https://zeroproject.org/practice/austria-signtime/

3.8 Writing through technology

The writing process has long benefitted from technologies that augment an author's writing skills. Examples of this technology include spelling and grammar checkers, automated translation of terms, computer-aided translation, automated compliance checks, or email writing assistance. Emailtree and Parlamind are examples of companies that offer the latter. The workflow of both companies is similar: an incoming email text is analysed by a language model, the system generates a response and a person responsible for this communication checks the text, does post-editings if needed, and sends it. Emailtree users report that they were not very proficient in writing in French, but they understand French very well. The software helped them write their email responses correctly. As a nice side effect, their French writing skills improved by looking at good examples. Both companies advertise that the email responses can be produced in seconds. The software does not replace human-human communication, but augments human writing with speed and L2 writing skills. These users are writing *through* technology, these facilities taking an active part in their language use.

Text processors, e.g. Microsoft Word, have long integrated spelling and later grammar checkers for an increasing number of languages. These were extremely helpful to spot the odd mistake. The technology evolved up to predictive writing systems, which are now available, both commercially and for free, across different systems and platforms. Computer text-processing tools can now integrate applications such as Grammarly or Language Tool, which help writers produce their text, in much the same way as some cloud-based services, such as Google Docs or mobile devices. Some of these tools are designed for academic and professional settings, while others were designed specifically to assist second language acquisition (SLA). Specific grammar checkers to support SLA have been increasingly developed (e.g., Felice 2016; Blazquez 2019), including for smaller languages (e.g., Estonian), so as to help students develop their own writing and receive corrective feedback on their free writing. This is the case of the free online resource UNED Grammar Checker (http://portal.uned.es/portal/page?_pageid=93,53610795&_dad=portal&_schema=PORTAL).

Currently, systems like Apple's predictive writing or Google Docs guess what one's next word is based on previous words, which allows writers to speed up typing, find that 'missing' word, and make fewer mistakes. The improvement of the technology over time inevitably had an impact on an author's writing: one can speculate that as more writing assistants are available, and of a higher quality, the line between human and machine in the writing process will blur. This blurring is precisely a defining feature of the human-machine era.

Other tools use the results of the linguistic and NLP analysis of large corpora (by learning, for instance, the use of typical collocations or multi-word units) to help users improve the vocabulary range, specificity and fluency of their academic English writing. A tool like Collocaid, for example (https://collocaid.uk/prototype/editor/public/), does so by suggesting typical collocations for a large number of academic terms. Users can start writing directly in the ColloCaid editor, or paste an existing draft into it.

Textio Flow (https://textio.com) advertises itself as an augmented writing tool that can 'transform a simple idea of a few words into a full expression of thoughts consisting of several sentences or even a whole paragraph'. The system, then, goes farther than simple text-processing tools, because they allow text to be written from a conceptualisation of the writer's ideas. See also: https://www.themarketingscope.com/augmented-writing/

In professional settings, writing assistants have been used, for example, to mitigate gender bias in job advertisements. Hodel et al. (2017) investigate whether and to what extent gender-fair language correlates with linguistic, cultural, and socioeconomic differences between countries with grammatically gendered languages. The authors conclude that a correlation does indeed exist, and that such elements contribute to socially reproduce gender (in) equalities and gender stereotypes. Similarly, Gaucher et al. (2011) investigate whether gendered wording (which includes, e.g. male-/ female-themed words used to establish gender stereotypes) may contribute to the maintenance

of institutional-level inequality. This is especially the case of gendered wording commonly used in job recruitment materials, especially in roles that are traditionally male-dominated. Several tools can be used to balance such gender bias in job advertisements. An example is Textio (https://textio.com), a tool that seeks to provide more inclusive language, which assesses a preexisting job description, scores it and subsequently makes suggestions on how to improve the writing and obtain a higher score - which means more applications, including from people who otherwise would not apply. Unlike Textio, which is a subscription-based tool, Gender Decoder (http://gender-decoder.katmatfield.com) is a free tool that assists companies, by reviewing their job descriptions. The tool makes suggestions based on a word list to remove words associated with masculine roles, hence discarding gender bias.

However, gender bias is not exclusive of corporations. Recent research has found that court decisions tend to be often gender-biased, despite the expectations that the law treats everyone equally. Recent research (Pinto et al. 2020) has thus focused on building a linguistic model that will be used to develop a writing assistant that will be used by legal practitioners. The tool will flag the drafted text for possible instances of gendered language and subsequently draw the attention of the writer to those instances.

3.8.1 Professional translators as co-creators with machines

Many professional translators prefer translation software that supports the translator's work by leveraging Translation Memory (TM); this uses deep learning to predict their inputs, which reduces certain repetitive aspects of their work (see e.g. Zhang et al. 2020), also equipped with domain-specific specialist terminology and text formatting. These packages are collectively known as CAT (computer-assisted translation) tools. An example of a tool for specific terminology is IATE (Interactive Terminology for Europe, https://iate.europa.eu/), the EU's terminology database. IATE provides users with accurate translations of institutional terminology in all EU working languages. IATE relies on a static database, not machine learning; and this speaks precisely to the limitation of current machine learning, that it is not entirely accurate and precise when it comes to professional or institutional terminology. Machine translation is based on probabilistic models, essentially educated guesses about which word equates to which other word in another language, according to similar pairs in the training corpus. That is currently not accurate enough for institutional needs; and so it remains useful to combine them with CAT tools.

Nevertheless, machine translation of written text increasingly complements CAT tools. Some of the largest translation vendors worldwide have long enabled translators to connect their CAT tool to machine translation engines like Google Translate. This allows translators to auto-translate a text segment if it does not exist in the TM. More recently, translation companies have developed proprietary machine translation systems. One of the main advantages of these is that access is controlled, and hence the risk of violating copyright is smaller. An example of such technology is the RWS Language Cloud (formerly known as SDL Language Cloud). The role of the human translator has thus shifted from a quiet, independent human working at home, to a 'techo-translator', trained to use technology for their own advantage, rather than resist it. This is one of the main applications of machine translation, but not the only one.

The issue of privacy in CAT has long been a topic of concern for translators. The fact that TM can be used to speed the translation process (while lowering the translators' fees) has raised concerns over intellectual property. If several translators contribute to a translation project, it becomes unclear whose property is the TM.Add to this the fact that translations usually include confidential information that may be compromised when using translation systems, in general, and MT in particular. MT has multiplied these concerns exponentially, as identifying the contributors becomes virtually impossible. In any future developments, therefore, it will be important to embed privacy as a core element.

But nowadays most of the machine translation in the world is done for non-professionals: for people living in a multilingual area, travelling abroad, shopping online in another language, and so on. Machine translation is not yet fully accurate but usually good enough for the gist of text or speech in supported language pairs. is also used on a daily basis by monolingual speakers - or writers - of a language, or by people who can't speak a certain language, to get a gist of the text or to produce texts immediately and cost-effectively where hiring a human translator is not possible. The result cannot yet be compared to a human-translated text, but it allows access to a text that would otherwise be impossible.

Machine translation can also be used at a more technical level, as part of a process to further train and refine machine-translation systems. Texts can be auto-translated, then 'post-edited' (corrected) by a human translator, then this output is used to train new machine translation systems.

Machine translation can also be used to perform other tasks. An example of such an application is translingual plagiarism detection (Sousa-Silva 2014), i.e. to detect plagiarism where the plagiarist copies the text from a source language, translates it to another target language and uses it as their own. The underlying assumption is that because plagiarism often results from time pressure or laziness, plagiarists use machine (rather than human) translation. Hence, by reverting the procedure - i.e. by translating the suspect text back - investigators can establish a comparison, identify the original and demonstrate the theft.

3.8.2 Cyber Journalism, Essay and Contract Generation

OpenAI is a company founded in 2015 by Elon Musk and other tech gurus. In 2019, they announced a revolutionary AI system, 'Generative Pre-trained Transformer 2' (or GPT-2). Based on deep learning of vast databases of human texts, GPT-2 was able to produce news stories and works of fiction independently, ~~based on~~ only short initial prompts, based on predictions of what should come next. The system was then dubbed "deepfakes for text". The quality of the text automatically generated by GPT was reported to be so high that OpenAI refused to publicly release it before the possible implications of the technology were discussed, on the grounds that the risk of malicious use was high. Among the negative implications identified by OpenAI are openly positive or negative product reviews, for example, or even fake news and – because it is trained using Internet texts – hyperpartisan text and conspiracy theories.

Its successor, GPT-3, was announced in 2020. It was trained with 175 billion parameters, which allowed it to produce news stories, short essays, and even op-eds. An example of an automatically generated piece of text is that of a Guardian news article written by GPT-3: https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3.

The potential of GPT-3 was reported by Brown et al. (2020), who tested the system over a set of tasks related to text generation, including (among others):

- Language modeling, cloze and completion tasks;
- Closed book question answering;
- Translation;
- Identification of a grammatically ambiguous pronoun referent;
- Common sense reasoning;
- Reading comprehension;
- Natural language inference.

Among the applications enabled by automatic text generation systems are:

- Cyber Journalism;
- Legal text/contract generation;
- Automatic essay writing.

As reported by Brown et al. (2020), despite the wide range of beneficial applications, the system still has several limitations. It tends to perform well when completing a specific task if trained over such a task; conversely, it tends to perform less efficiently when trained over a diverse range of materials to perform a diverse set of tasks. Indeed, the quality of the text currently generated automatically by AI tools is obviously largely dependent on the text genre. For example, smart document drafting is already used as a service provided by companies for instance to automatically draft contracts (see https://www.legito.com/US/en; https://www.erlang-solutions.com/blog/smart-contracts-how-to-deliver-automated-interoperability.html). However, the success of these servicse is reliant on the fact that this text genre is highly formulaic and controlled, and because often only some minor details change across different contracts. Automatic text generation can be less fruitful in other areas, or when producing other text genres.

Automatic text generation systems, like GPT-3, do not have a reasoning of their own; hence, they do what computers have done in recent decades: augment the authors' writing experience. Additionally, because they are trained - at least for now - on text produced by humans, they tend to carry with them some major human flaws, including bias, fairness and skewed representation (Brown et al. 2020).

Another challenge faced by automatic text generation systems is the human skills of recursion and productivity.

Recursion is the human property that allows humans to incorporate existing sentences and sentence excerpts in similar linguistic frames; productivity is the property that allows speakers and writers of a language to combine a few linguistic elements and structures to generate and understand an infinite number of sentences. Therefore, given a relatively small number of elements and rules, "humans can produce and understand a limitless number of sentences" (Finegan 2008) - which systems, even those as sophisticated as GPT-3, still struggle to do.

However, regardless of whether GPT-3 (or a successor, GPT-*n*) can produce text independently, or whether it is used to augment the writing process, the implications will be significant. Think for example about automatic essay generation. One of the basic tenets of academic and scientific writing is integrity and honesty. Currently, 'cheating' students may resort to 'essay mills', private companies or freelancers who write essays for money. Although current technology is somewhat limited, future developments could bring high quality essays into the purview of AI, including accurate approximation of the student's own writing style (assuming they have previously written enough themselves to train the AI) (Feng et al. 2018). But by the same token, that same AI could detect whether an essay had been created by AI. We may be in for something of a virtual arms race.

Although the system performs better than previous machine learning systems in some of these tasks, it is not yet on a par with human-generated text. Yet.

3.9 Personality profiling

The advance of language technology opens up new opportunities for many interdisciplinary fields, for example text analysis and psychology. Personality profiling is used in many tasks such as recruiting, project facilitation, and career development. Traditionally, personality profiling is done by designing a series of questions with each pointing to a predefined scale. The scale normally ranges from low to high concerning the agreement on the statement of the question.

A more flexible approach can be applied with natural language processing and machine learning. The assumption is that the answers written in a 'free form' to replace the predefined, scaled answers are more difficult to manipulate, and more sensitive to individual variation. To leverage this assumption, open questions can be used to lead the participants to write free text answers. Since the construct of language models has reached far beyond mere domain vocabulary-matching and word-counting, a deeper understanding of the text considering the context and implicit information can be expected. In the case of personality profiling, a model can be trained to capture not only the content of the text but also the way the information is conveyed. A real-life application of such a solution is implemented at Zortify (zortify.com), which claims to enable participants to express themselves freely in a non-competitive environment. A large amount of annotated, multilingual data is collected for training a deep learning model: 30 thousand combinations of personality-related, open questions. It has been found that the length of the text influences the performance of the model: the analysis is more accurate when the answer is relatively long. In addition, it undermines the performance when the answers are not related to the questions, which is as expected.

There are multiple and overlapping concerns about the use of automated personality profiling. Traditional models work well under certain constraints while performing poorly when the constraints cannot be met. Imagine a scenario where the questionnaires are applied to participants in a competitive environment, the answers to the questions can be manipulated such that a 'perfect' profile is created to oversell without revealing the real personality of the participant. Questions are also clearly based on certain cultural assumptions about behaviour, values, and other subjective traits. Profiling is at the very least a simplistic and reductive exercise in categorising humans; at worse it can be racially discriminatory or otherwise unappreciative of cultural differences (see Emre 2018). Involving AI could introduce some nuance into this, but could also add new problems. Concerns a rise about data security, since a lot of personal data must be handed over for these tests to operate; and relatedly there is the issue of consent - if you refuse to surrender your data and therefore fail the interview. Moreover, any profiling, however intelligent, is based on a snapshot and will struggle to handle changes in individual humans' lives, while also potentially screening out those with mental health or other conditions that may present as undesirable personality traits. Further, as machines become superficially better at performing a given task, they can be trusted more fully and unquestioningly by those who employ them. This can further reduce the scope for appreciating nuance (Emre 2018). Progress in this area will clearly centre on issues of trust, consent, equality, diversity, and discrimination, many of which - as discussed earlier - can be baked into a system by virtue of biases in the training data.

4 Language in the Human-machine Era

4.1 Multilingual everything

We have outlined a near future of ubiquitous, fast or even real-time translation. This will bring new issues of user engagement, trust, understanding, and critical awareness. There is a need for "machine translation literacy" to provide people with the critical resources that allow them to become informed users of the tools available. One relevant project on this is the Machine Translation Literacy project (see, e.g., Bowker 2020).

The technological developments of the last decades have given rise to a "feedback loop": technology influencing our production of language. It is not uncommon (especially for large companies) to write *for* machine translation. This process, which is known as pre-editing, consists of drafting a text following a set of rules that keep in mind that the produced text could be easily translated by an MT engine and require a minimal post-editing effort, if any. This involves, e.g., mimicking the syntax of the target language, producing simple and short sentences, avoiding idioms and cultural references, using repetitions, and so on. This is not a new procedure, since controlled language has been used for decades, especially in technical writing, to encourage accuracy and technical/scientific rigour. More recently, language technology companies have encouraged their content creators to use such controlled language, and this can ultimately lead to a simplified use of language (see, e.g., IBM, United Language Group, or Ajujaht). Meanwhile for the user of technology, similar issues may arise; see for example Cohn et al. (2021) showing the way people may adjust their speech patterns when talking to a virtual assistant. If we spend more time talking *to* technology, this effect may grow. In due course, with enough exposure, it could lead to larger changes in human speech, at least amongst those most frequently connected. All this is wide open space for research in the human-machine era.

4.2 The future of language norms

Changes in media use generally make the medium 'visible' (see McLuhan 1962). Earlier media of writing and print influenced our ideas about language, but also memory culture and social order (see e.g. Ong 1982, Assmann 2010). Norms of spelling and grammar have emerged from socio-historical developments, in line with histories of technologies of writing and print. If the human-machine era increases our awareness of the role of machines, it may move us to see language more as a code, an algorithm, a denaturalised sterile entity. This could have all manner of outcomes, from a decrease of standard language as a social authority, through to the foundational concepts of languages. Perhaps a more frequent connection of language *to* technology - inherently indifferent to national boundaries - will loosen the bonds of linguistic nationalism that have so guided popular understandings of language since the middle ages.

At the same time, traditional national norms of language are often programmed into digital devices and thus stabilized in digital uses. To what extent AI tools, which typically react to statistical frequency, will produce new norms, not based on what was formerly thought of as 'correct' but on what is frequent, remains to be seen. Linguistics itself is of course a strictly descriptive science, uninvested in value judgements (see e.g. Schneider 2020, about what is correct language). A key area for future investigation will be how technology could contour our judgments on what kinds of language are correct, authentic, grammatical, and otherwise acceptable.

As informal written communication becomes increasingly mediated by speech-to-text software, will we guide or be guided by the auto-completed sentences provided to us by technology? Current trends in the use of machine-mediated human communication are likely to result in various changes to the perception and role of (at least) spelling in society; the default rules used by spell checkers, text-to-speech software, etc. will play an important part in this process, to an extent that individual authorship, too, will change as a consequence.

Even the predictive typing tools such as the Google Doc's Smart Compose feature have an influence on how we express our ideas in language - and how one learns with them. Because such systems learn from rules and usage statistics, automated suggestions tend to reduce text diversity (see e.g. Arnold et al. 2020). Hence, tools like *Grammarly* or *Language Tool* may contribute to less diversity over time.

4.3 Natural language as an API

Natural, human language is increasingly used as an application programming interface (API). Conventionally APIs are integrated into existing systems. Using natural language as an API allows AI systems to communicate with each other and learn from each other. An example of this in action is the use of Google Assistant: although this is a novel crowdsourcing approach, the system was able to extract information from an automated call center AI (https://www.polyai.com/our-voice-assistant-spoke-to-google-duplex-heres-what-happened/). From populating Google Maps with a business's opening hours, to interacting with other voice command units, the possibilities are immense.

In the coming years, this technology will likely be made available to many of us; we will be able, for example, to give a voice instruction to our smartphone assistant, which will then be able to contact different airlines to book us a flight ticket, or check with hotel receptions that they have a free room, and on and on into a virtually blurred world of human-machine ease, complexity, simplicity, challenge, freedom, and constraint.

4.4 Problems to solve before using NLP tools

The use of NLP tools does not come without problems, one of which is bias (Blodgett et al. 2020). Issues like these are not surprising since those tools build upon natural language, produced by real people, under common circumstances. Discrimination is endemic. It is rife in the data. The blame lies not with the machines, but with us. They are built in our image. But identifying this issue is the beginning of addressing it in research, some early stages of which we have covered in this report.

As they have social and policy goals for equity and fairness in mind, researchers in the field have recently recentered their focus on studying the statistical properties required to achieve language models that are fair (Saleiro

et al. 2020). In fact, the awareness of biases and the availability of potential automatic filtering devices or more fine-grained statistically "fair" models will contribute to feeding machine learning with more representative and inclusive training data, which will find its reflection in output that is sensitive towards bias and discrimination.

The following questions are likely to be at the centre of research at least in the coming years: How will diversity be represented? Synthetic voices will be able to mimic a wide array of accents and dialects, as we have noted so far, and so to what extent will we choose difference over sameness in talking to and *through* technology? Could all this have a positive impact on society, by helping avoid linguistic discrimination on the basis of accent? Or, in contrast, with AI tools that are programmed to detect frequent patterns, will there be even less diversity?

Studies in other realms (e.g. image recognition, face recognition) have shown severe problems, as in minority groups or women being misrecognized, due to lower amounts of data with these groups. This makes it all the more pressing to accelerate efforts to develop machine-readable corpora for minority languages. The human-machine era is coming. We will be interacting evermore through and *to* technology. Some languages and language varieties will be excluded, at least to begin with. The research community must rise to this task of equalising access and representation.

A number of initiatives have arisen in the past 10 years on ethical and legal issues in NLP (e.g. Hovy et al. 2017, Axelrod et al. 2019). An example of an ethical application of NLP is the deployment of tools to identify biased language in court decisions, thereby giving the courts an opportunity to rephrase those decisions (Pinto et al. 2020). Another is in the detection of fake news. Until now, most computational systems have approached fake news as a problem of untruthful facts, and consequently have focused on fact-checking or pattern recognition of fake news spreading. The problem, however, is that misinformation is not necessarily based on true or false facts, but rather on the untruthful representation of such facts (Sousa-Silva 2019). Hence, more sophisticated systems are required that are able to identify linguistic patterns to flag potential fake news. A team of researchers, building upon previous work (Cruz et al. 2019) that focused on hyperpartisan news detection, is currently developing such a system. This is another area to watch for progress in the coming years.

4.5 Speechless speech

In 2019, the first multi-person non-invasive direct brain-to-brain interface (BBI) for collaborative problem solving was introduced: BrainNet (Jiang et al. 2019). BBIs enable communication between two brains without the peripheral nervous system. They consist of two components: a brain-computer interface (BCI) that detects neural signals from one brain and translates them to computer commands; and a computer-brain interface (CBI) that delivers computer commands to another brain. The first experiment with BrainNet was quite simple. It involved flashes of light to communicate simple responses like 'yes' and 'no'. However, R&D in this field aspires to achieve full conversations between minds. This may seem daunting, but has clear applications first in healthcare (among paralysed people or those with locked-in syndrome), and later foreseeably among workers in locations where audible conversation is impossible (loud noise, in space, underwater). Future application in consumer devices could quite foreseeably follow.

More recently, MIT researchers have developed a computer interface that transcribes words that the user verbalizes internally but does not actually speak aloud. This is known as subvocalization detection: it consists of using wearable and implanted technology to allow for such speechless conversations. The system developed at MIT consists of a wearable device and a computing system. Electrodes in the device pick up neuromuscular signals in the jaw and face. The signals are triggered by internal verbalizations but cannot be detected by the human eye. They are fed to a machine-learning system that correlates particular signals with particular words.

We debated whether to include this kind of 'speechless speech' in the body of this report, or whether to leave it here as something of a postscript - as we did eventually. The purpose of this report is to describe language technologies that are almost within reach, very close to release and widespread consumer adoption. Products like Facebook's Aria and Oculus Rift already have the beginnings of a marketing machine building up hype. But despite the somewhat eccentric noises from Elon Musk about his *Neuralink* brain interfaces, nevertheless the technology is still palpably further from widespread adoption. That said, just as with AR and VR so too with BBIs there is significant corporate investment. We will return to this in subsequent editions of this report. Perhaps it will be upgraded into the main body of the report in years to come.

Moreover, BBIs present the potential for transformations that are magnitudes beyond the comparatively trivial tweaks represented by the AR and VR gadgets reviewed in this report. The potential for communicating directly between our brains could ultimately enable communication without language, or with some mix of language and inter-brain fusion that we cannot begin to imagine yet. But with great power comes great responsibility. There will come great risks, not only of surgery and physical brain damage, but also risks to autonomy, privacy, agency, accountability and ultimately our very identity (Hildt 2019).

4.6 Artificial Companions and AI-Love

Research shows that some users of voice-controlled devices develop social and emotional relationships with them (see e.g. Schneider, forthcoming), which has an enormous potential in human feelings such as love (Zhou & Fischer 2019). In a similar vein, social robot companions with different built-in services are getting evermore sophisticated. Robot companions can be used to help autistic children develop social skills, empowering vulnerable individuals (Damiano & Dumouchel 2018).

Something we have not covered in much detail so far is the possibility for AR and VR to be used for sexual gratification (Damiano & Dumouchel 2018; Coursey et al. 2019; Zhou 2019). There may be nothing necessarily or inherently wrong with 'sex robots'. However, such a technology could potentially be used to synthesise immoral or illegal sex acts (Damiano & Dumouchel 2018). This is (thankfully) beyond the concern of a report focused on language, but will be a very real topic of debate brought to life by the same technologies we have outlined so far.

Moreover, language in the human-machine era cannot be approached without a discussion of societal, moral, ethical and legal issues that will inevitably arise. Since all technologies, regardless of their intrinsic value, can be used for unjust ends, it is an obligation to build in an understanding of - and safeguards against - such harms at the design stage.

4.7 Privacy of linguistic data and machine learning

Individuals and organizations across the world may be increasingly privacy-conscious, even if paradoxically they tend to grant more access to their data. While users are increasingly made more aware of their data being accessed, they are made aware of their right to privacy, but also informed that refusal to surrender data means exclusion from a given product or service. In some cases the 'price' of our data is made explicit by differential pricing for 'enterprise' customers who typically enjoy stricter controls. Google for example has a reassuring page for business clients on data protection: https://business.safety.google/compliance/. Regular users of its free services meanwhile are told that their data will be mined for advertising and other purposes. Facebook's Oculus Rift VR headset is available to consumers for $299, in a format that only works when connected to a Facebook account - harvesting data including "your physical dimension, including your hand size, .. data on any content you create using the Quest 2, as well as … your device ID and IP address" (Dexter 2021). However, the enterprise option with no Facebook requirement is available for €799, plus an annual service fee of $180 (ibid.). This side-by-side comparison lays bare the price of 'free', and the value of your data. As new intelligent devices collect inestimably more data than the current rudimentary harvesting of clicks, taps and swipes, so too the debate will evolve over privacy, security, and private control of personal information.

4.8 Decolonising Speech and Language Technology

Language colonisation (and consequently the need for decolonisation) has long been discussed (see e.g. Williams 1991). The colonial era saw indigenous and non-standard language varieties variously displaced, eradicated, and often banned. Colonial languages like English, Spanish, Portuguese and French became institutionalized as official languages of communication, and an instrument of political and ideological domination (Juan 2007, Macedo 2019).

Language decolonisation has traditionally targeted foreign and second language education that overlooks 'regional' languages while foregrounding western thought and language (Macedo 2019). More recent research has focused on

colonising discourses in speech and language technology. For example, Bird (2020) invites us towards a postcolonial approach to computational methods for supporting language vitality, by suggesting new ways of working with indigenous communities. This includes the ability to handle regional or non-standard languages/language varieties as (predominantly) oral, emergent, untranslatable and tightly attached to a place, rather than as a communication tool whose discourse may be readily available for processing (ibid.).

This approach to technology has attracted the interest of linguists and non-linguists alike. An example of the latter is that of Rudo Kemper, a human geographer with a background in archives and international administration, and a lifelong technology tinkerer, whose work has revolved around co-creating and using technology to support marginalized communities in defending their right to self-determination and representation, thereby contributing towards achieving decolonizing and emancipatory ends. This is the case, in particular, of his work with Digital Democracy on the programs team, where he leads the creation of the Earth Defenders Toolkit (https://earthdefenderstoolkit.com). The aim of this project is to provide communities with documents, tools and materials that foster community-based independence. To that end, the project supports communities' capacity-building, local autonomy, and ownership of data.

4.9 Authorship and Intellectual Property

As text-generation technologies evolve, the potential of technological applications to improve human-machine interaction is enormous. In language learning and teaching, students can learn from and with technology. Streamlined and more efficient writing assistants will allow both native and non-native speakers of a language to produce more proficient texts, more efficiently (though, as we have cautioned, at different rates of development for different languages). Inevitably, however, using language is exerting power (Fairclough 1989) and computers will shape the ideas of people to a greater or lesser extent. The work of Fairclough and other discourse analysts has centred in the way power is embedded in language. As machines increasingly influence and co-create our language, technology will clearly become unavoidable for future analyses of discourse and power.

Technology will also affect authorship, as it may be used for fraudulent practice. One of the current concerns in education institutions around the world is 'contract cheating' (e.g. https://www.qaa.ac.uk/docs/qaa/guidance/contracting-to-cheat-in-higher-education-2nd-edition.pdf). Future AI is likely to be able to produce high quality essays, as we discussed earlier (see Feng et al. 2018). There are issues here for both plagiarism and its detection, which will come to the fore in educational settings in the coming years.

With regard to intellectual property, if we are talking and writing *through* technology - the tech is a genuine co-creator of our language - the question emerges 'who is the author?' and 'who owns the copyright?'. These have been topics of debate among translators, translation companies and their clients, but they will span far beyond the translation industry. In most jurisdictions, the rights of authors are protected both in their moral and in their financial dimensions (Pereira 2003). The latter grants authors the right to financially explore their intellectual property, whereas the former entitles them to enjoy the integrity of their creative production. Thus, as machine learning systems feed upon the texts available by crawling the web, both those rights can be difficult to guarantee due to the fragmentary and especially the ubiquitous nature of the data, which can make it difficult even for authors to be aware of the reuse of their texts. In this context, even the concept of 'fair use' may need to be redefined.

4.10 Affective chatbots

Attitudes towards conversing with technology are changing with advances in conversational agents. In ordinary, daily interaction, both pragmatic attributes and the ability to entertain, show wittiness and tell jokes (like humans) are valued in chatbots (Følstad & Brandtzaeg 2020). The question regarding user information is: when, how and to what extent, if at all, should humans be informed that they are interacting with a machine?

Some studies show that humans tend to align/entrain their voices and style of speaking to the automatic/robotic conversational partners (see e.g. Bell et al. 2003; Sinclair et al. 2019; Zellou & Cohn 2020). Previous research has shown that humans have a natural tendency to both anthropomorphise ands adjust our language and behaviour according to expectations (Pearson et al. 2006; Nass 2004; Nass & Yen 2010). A popular debate coming soon will be whether it matters or not to be talking to other humans, especially if one already perceives the technology as

another being; and whether - and at what stage - we will be speaking *to* technology in the way that we speak to other humans. As the lines separating chatbots from humans become increasingly blurred, new affective robots and chatbots bring a new dimension to interaction, and could become a means of influencing individuals. The ethical issues underlying affective computing are myriad (see Devillers et al. 2020 for an extensive discussion).

4.11 Ethics, lobbying, legislation, regulation

Lobbying has been both a topic of concern and discussion over the years (see for example https://www.lobbycontrol.de/schwerpunkt/macht-der-digitalkonzerne/), with varying arguments as to its morality, legality, and place in civil society. Naturally this discussion gravitates towards regulation and transparency. According to the German independent organisation LobbyControl (https://www.lobbycontrol.de/2021/01/big-tech-rekordausgaben-fuer-lobbyarbeit-in-den-usa/), big tech companies are among the most active lobbyists in the world. In the European Union, legislators attempted to limit their powers of influence by passing the *Digital Services Act* (https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en), which "aims to ensure a safe and accountable online environment" and the *Digital Markets Ac*t for "fair and open digital markets" (https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en). These two laws can have a significant impact on how big tech companies perform their activities in Europe, including on language models.

The 'good'/'value-added' use of technologies in language learning and teaching circles poses numerous challenges. One key challenge (that must be addressed in pre- and in-service teacher training) is the smooth articulation of technology, pedagogy and content knowledge, or TPACK, as put forward by ~~Mishra & Koehler (2006 and 2014)~~. The three main components of TPACK in language learning and teaching (as presented in Meunier 2020) are: content knowledge (CK), defined as the teachers' knowledge about the subject matter to be learned or taught (an additional language here); pedagogical knowledge (PK), which is the teachers' knowledge about the appropriate and up-to-date processes and practices or methods of teaching and learning additional languages; and technological knowledge (TK), which consists of an in-depth understanding of technology, enabling the teacher to apply technology productively to assess whether it can assist or impede the achievement of a learning goal. This integration is not guaranteed by the sheer existence of technology and it must constantly be kept in mind by regulators.

A further example of regulation impacting technology is language generation models, such as GPT-3 discussed earlier, which builds on huge volumes of data. Hutson (2021) discusses the tradeoffs between the interests of big technology companies to make the models increasingly larger, and the need to make those models safer and less harmful to non-dominant groups. One of the concerns raised by the author is that researchers working for big tech companies may face serious problems with their employers because of their concerns with the ethical implications of big language models (Hutson 2021). These facts justify the need for legal regulation of the language technology market, such as the one drafted by the European Union, for a more ethical human-machine era.

It seems clear, however, that such regulation and legislation should not be accomplished at all cost; rather, legislation and technology must be balanced so as not to show research and development in the areas, or even stop technological progress and innovation.

Another (probably slightly longer-term) issue will be the role of teachers and the profound changes that are likely to take place in the teaching profession. If chatbots can train learners with basic conversational skills, should teachers go on teaching speaking skills? If augmented automatic translation is available, should we even go on teaching languages? And if so, which ones? Should teaching be refocused on critical literacy (by, for instance, making sure that learners/users can critically evaluate translations to avoid being manipulated)? If those questions are likely to impact teaching and learning circles in ways that are probably not conceivable today, we should start thinking about these, now. Forewarned is forearmed.

5 Conclusion, looking ahead

At this particular moment in human history, it may feel somewhat of a dangerous distraction to spend so much time thinking about gadgets for watching movies, playing games, and taking the effort out of simply talking to each other. There are, it must be said, slightly more pressing issues afoot, as the following tweet wryly puts it:

Will be re-screenshotted

https://twitter.com/etienneshrdlu/status/1388589083765710854

We are not here to cheerlead for new language technologies, or to suggest they are anywhere near as important as the future of the planet, or other issues like access to clean water, healthcare, and so on. Rather, we begin from the position that these technologies are coming, thanks to the huge and competing private investment fuelling rapid progress; and that we can either understand and foresee their effects, or be taken by surprise and spend our time trying to catch up.

Debates on Beneficial and Humane AI (https://humane-ai.eu) may be a source of inspiration for debate on new and emerging language technologies. Dialogue will enrich and energise all sides - see https://futureoflife.org/2017/12/27/research-for-beneficial-artificial-intelligence/, https://uva.nl/en/research/research-at-the-uva/artificial-intelligence/artificial-intelligence.

This report of ours has so far sketched out some transformative new technologies that are likely to fundamentally change our use of language. Widespread AR will soon augment our conversations in real time, while showing us information about the world around us. We will be able to talk in noisy or dark environments easily, and across multiple languages - as voices and faces are augmented to sound and look like they are producing other languages. We will be able to immerse ourselves in virtual worlds, and interact with a limitless selection of lifelike virtual characters equipped with new chatbot technology able to hold complex conversations, discuss plans, and even teach us new things and help us practise other languages.

Some of these may feel unrealistically futuristic or far-fetched, but a central purpose of this report - and the wider LITHME network - is to illustrate that these are mostly just the logical development and maturation of technologies currently in prototype. Huge levels of corporate investment lie behind these new technologies; and once they are released, huge levels of marketing will follow.

One of the founders of RASA (open-source chatbot development platform), Alan Nichol, predicted in 2018 that chatbots are just the beginning of AI assistants in the enterprise sector. His forecast on the future of automation in the enterprise sector is structured into five levels: automated notifications, FAQ chatbots, contextual assistants, personalised assistants and autonomous organisations of assistants. If his prediction were true, we must be in the age of contextual and personalised assistants as we write, in 2021. Autonomous organisations of assistants - to appear in 5-7 years, according to Nichol's forecast - or *multi-agent system*, will leave the academic niche and reach the "normal" users. See for instance the AAMAS conference for an overview of topics and problems: https://aamas2021.soton.ac.uk/.

But will everyone benefit from all these shiny new gadgets? Throughout this report we have tried to emphasise a range of groups who will be disadvantaged. This begins with the most obvious, that new technologies are always out of reach for those with the lowest means; and the furious pace of consumer tech guarantees the latest versions will always be out of reach. As these new technologies mature and spread, this may evolve into an issue of government concern and philanthropy. It may seem fanciful to imagine that futuristic AR eyepieces could become the subject of benevolence or aid; but there is recent historical form here. Only two decades ago, broadband internet was a somewhat exclusive luxury; but today it is the subject of large-scale government subsidy (e.g. in the UK: https://gov.uk/guidance/building-digital-uk), as well as philanthropic investment in poorer countries (https://connectivity.fb.com/, https://www.gatesfoundation.org/our-work/programs/global-development/global-libraries); and even a UN resolution (https://article19.org/data/files/Internet_Statement_Adopted.pdf). In due course, the technologies we discuss in this report could go the same way. VR for all.

Then there are less universal but still pressing and persistent issues of inequality, which we have also highlighted. Artificial intelligence feeds on data. It churns through huge datasets to arrive at what are essentially informed guesses about what a human would write or say. More data equals better guesses. Little or no data? Not so useful. The world's bigger languages - English, Mandarin, Spanish, Russian, and so on - have significant datasets for AI to chew over. For smaller languages, this is a taller order. Progress in 'transfer learning', as we reviewed in this report, may help here.

The inequality faced by minority languages is essentially just a question of gathering enough data. The data in question will be in exactly the same format as for bigger languages: audio recordings, automatically transcribed and perhaps tidied up by humans for AI to digest. But for sign languages, there is a much bigger hill to climb. Sign language is *multimodal*: it involves not only making shapes with the hands; it also relies heavily on facial expression, gesture, gaze, and a knowledge of the other signer's own life. All that represents a much greater technological challenge than teaching a machine to hear us and speak like us. For the Deaf community, the human-machine era is less promising.

Important issues of security and privacy will accompany new language technologies. AR glasses that see everything you see, VR headsets that track your every movement; these devices will have unprecedented access to incredibly personal data. Privacy policies for technologies are not an open buffet. You either accept all the terms, or no gadget for you. This is playing out in the news at time of writing, with a controversial update to Facebook's instant messenger WhatsApp, enabling Facebook to "share payment and transaction data in order to help them better target ads" (Singh 2021). The two choices are to either accept this or stop using WhatsApp. Reportedly millions are choosing the latter. But this is actually a fairly minor level of data collection compared to what lies ahead with AR and VR devices, able to collect magnitudes more data. When those companies try to sneak clauses into their privacy policies to monetise that data, we may look back whimsically at today's WhatsApp controversy as amusingly trivial.

A further caution to end with is to re-emphasise the current limitations of AI. It is very popular to compare the astonishing abilities of AI to the human brain; but as we noted earlier this is over-simplistic (see e.g. Epstein 2016; Cobb 2020; Marincat 2020). And as the technology progresses further, we may look back at this comparison with a wry smile. There is indeed a history of comparing the brain to the latest marvellous technology of the day, which has inspired some gentle mockery here and there, for example the following tweet:

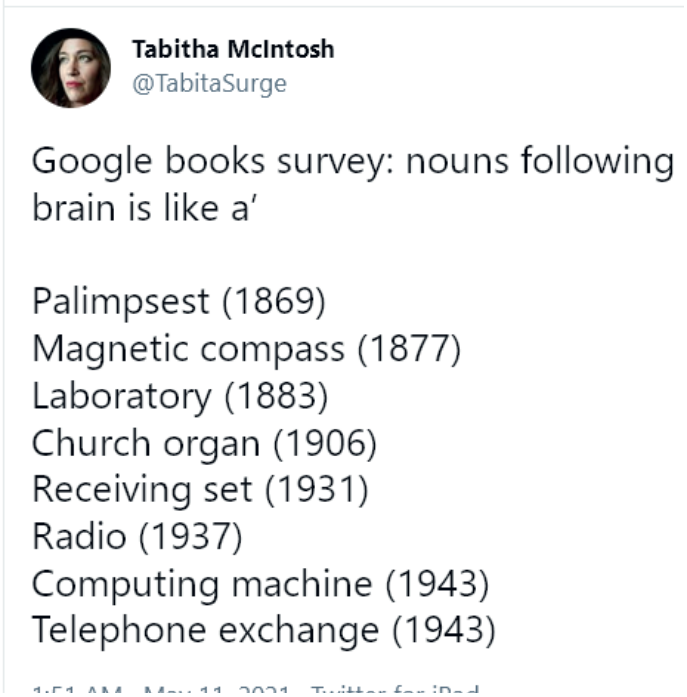


Figure XX: https://twitter.com/TabitaSurge/status/1391888769322831878

The strong reliance of AI on form, rather than meaning, limit its understanding of natural language (Bender & Koller 2020). In making meaning, we humans connect our conversations to previous conversations, common knowledge, and other social context. AI systems have little access to any of that. To become really intelligent, to reason in real-time scenarios on natural language statements and conversations, even to approximate emotions, AI systems will need access to many more layers of semantics, discourse, and pragmatics (Pareja-Lora et al. 2020).

Looking ahead, we see many intriguing opportunities and new capabilities, but a range of other uncertainties and inequalities. New devices will enable new ways to talk, to translate, to remember, and to learn. But advances in technology will reproduce existing inequalities among those who cannot afford these devices, among the world's smaller languages, and especially for sign language. Debates over privacy and security will flare and crackle with every new immersive gadget. We will move together into this curious new world with a mix of excitement and apprehension - reacting, debating, sharing and disagreeing as we always do.

Plug in, as the human-machine era dawns.

Acknowledgements

First and foremost we thank our funders, the COST Association and European Commission (https://cost.eu/). Without this funding, there would be no report. We have further benefited from excellent administrative support at our host institution, the University of Jyväskylä, Finland, especially the allocated project manager Hanna Pöyliö. All the authors express their gratitude to supportive friends, family, and pets. Appropriately for a report like this, we should express satisfaction with Google Docs and Google Meet, which together enabled us to collaborate smoothly in real time, and avoid the horror of email attachments and endless confusion over the latest version. We recommend these two facilities for any such collaboration.

References

Abbasi, A., Chen, H., & Salem, A. (2008). Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums. *ACM Trans. Inf. Syst.*, *26*(3). https://doi.org/10.1145/1361684.1361685

Ackerman, R., & Goldsmith, M. (2011). Metacognitive regulation of text learning: On screen versus on paper. *Journal of Experimental Psychology. Applied*, *17*(1), 18–32. https://doi.org/10.1037/a0022086

Ahmad, W., Zhang, Z., Ma, X., Hovy, E., Chang, K.-W., & Peng, N. (2019). On difficulties of cross-lingual transfer with order differences: A case study on dependency parsing. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2440–2452.

Aharoni, R., Johnson, M., & Firat, O. (2019). Massively Multilingual Neural Machine Translation. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 3874–3884. https://doi.org/10.18653/v1/N19-1388

Ai, H. (2017). Providing graduated corrective feedback in an intelligent computer-assisted language learning environment. *ReCALL*, *29*(3), 313–334.

Ai, R., Krause, S., Kasper, W., Xu, F., & Uszkoreit, H. (2015). Semi-automatic Generation of Multiple-Choice Tests from Mentions of Semantic Relations. *Proceedings of the 2nd Workshop on Natural Language Processing Techniques for Educational Applications*, 26–33. https://doi.org/10.18653/v1/W15-4405

Aji, A. F., Bogoychev, N., Heafield, K., & Sennrich, R. (2020). In Neural Machine Translation, What Does Transfer Learning Transfer? In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 7701–7710). https://www.aclweb.org/anthology/2020.acl-main.688.pdf

Alonso, J. M., & Catala, A. (Eds.). (2021). Interactive Natural Language Technology for Explainable Artificial Intelligence. In *TAILOR Workshop 2020 Post-proceedings*. LNCS, Springer.

Amaral, L. (2011). Revisiting current paradigms in computer assisted language learning research and development. *Ilha Do Desterro A Journal of English Language, Literatures in English and Cultural Studies*, *0*. https://doi.org/10.5007/2175-8026.2011n60p365

Amaral, L. A., Meurers, D., & Ziai, R. (2011). Analyzing learner language: Towards a flexible natural language processing architecture for intelligent language tutors. *Computer Assisted Language Learning*, *24*, 1–16.

Amoia, M., Bretaudiere, T., Denis, R., Gardent, C., & Perez-beltrachini, L. (2012). A Serious Game for Second Language Acquisition in a Virtual Environment. *Journal on Systemics, Cybernetics and Informatics (JSCI*, 24–34.

Andresen, B., & van den Brink, K. (2013). *Multimedia in Education: Curriculum*. Unesco Institute for Information Technologies in Education.

Andrew, J. (2018). Human vs. Machine Translation. *Medium.Com*. https://medium.com/@andrewjames206/human-vs-machine-translation-fb41b28d01ff

Antaki, C., & Wilkinson, R. (2012). Conversation Analysis and the Study of Atypical Populations. In J. Sidnell

& T. Stivers (Eds.), *The Handbook of Conversation Analysis* (pp. 533–550). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118325001.ch26

Archer, D., Culpeper, J., & Davies, M. (2008). Pragmatic annotation. In M. Kytö & A. Lüdeling (Eds.), *Corpus Linguistics: An International Handbook* (pp. 613–642). Mouton de Gruyter.

Arnold, K. C., Chauncey, K., & Gajos, K. Z. (2020). Predictive Text Encourages Predictable Writing. *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 128–138. https://doi.org/10.1145/3377325.3377523

Artetxe, M., Labaka, G., & Agirre, E. (2018). *Unsupervised Statistical Machine Translation*.

Asif, M., Ishtiaq, A., Ahmad, H., Aljuaid, H., & Shah, J. (2020). Sentiment analysis of extremism in social media from textual information. *Telematics and Informatics*, *48*, 101345. https://doi.org/10.1016/j.tele.2020.101345

Asimov, I. (1941). *Three laws of robotics*. Runaround.

Assmann, A. (2010). *Erinnerungsräume. Formen und Wandlungen des kulturellen Gedächtnisses* (5th ed.). C.H. Beck.

Axelrod, A., Yang, D., Cunha, R., Shaikh, S., & Waseem, Z. (Eds.). (2019). *Proceedings of the 2019 Workshop on Widening NLP*. Association for Computational Linguistics. https://www.aclweb.org/anthology/W19-3600

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). *SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining*. 5.

Baevski, A., Zhou, H., Mohamed, A., & Auli, M. (2020). *wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations*.

Baron, N. S. (2013). Redefining Reading: The Impact of Digital Communication Media. *PMLA*, *128*(1), 193–200.

Barreira, J., Bessa, M., Pereira, L. C., Adão, T., Peres, E., & Magalhães, L. (2012). MOW: Augmented Reality game to learn words in different languages: Case study: Learning English names of animals in elementary school. *7th Iberian Conference on Information Systems and Technologies (CISTI 2012)*, 1–6.

Barriere, V., & Balahur, A. (2020). *Improving Sentiment Analysis over non-English Tweets using Multilingual Transformers and Automatic Translation for Data-Augmentation*.

Bell, L., Gustafson, J., & Heldner, M. (2003). Prosodic adaptation in human-computer interaction. *Proc. ICPhS, 3*, 833–836.

Bender, E. M. (2019). *A Typology of Ethical Risks in Language: Technology with an Eye Towards Where Transparent Documentation Can Help*. The Future of AI workshop. https://faculty.washington.edu/ebender/papers/Bender-A-Typology.pdf

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . *Conference on Fairness, Accountability, and Transparency (FAccT '21)*, 610–623. https://doi.org/10.1145/3442188.3445922

Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185–5198.https://doi.org/10.18653/v1/2020.acl-main.463

Benjamin, R. (2020). *2020 Vision: Reimagining the Default Settings of Technology & Society*. International Conference on Learning Representations. https://iclr.cc/virtual_2020/speaker_3.html

Beukeboom, C. J., & Burgers, C. (2017). Linguistic bias. In H. Giles & J. Harwood (Eds.), *Oxford Encyclopedia of Intergroup Communication*. Oxford University Press.

Bibauw, S., François, T., & Desmet, P. (2019). Discussing with a computer to practice a foreign language: Research synthesis and conceptual framework of dialogue-based CALL. *Computer Assisted Language Learning*, *32*(8), 827–877. https://doi.org/10.1080/09588221.2018.1535508

Biber, D., & Conrad, S. (2009). *Register, Genre, and Style*. Cambridge University Press.

Bigand, F., Prigent, E., & Braffort, A. (2020). Person Identification Based On Sign Language Motion: Insights From Human Perception and Computational Modeling. *Proceedings of the 7th International Conference on Movement and Computing (MOCO '20)*, *Article 3*, 1–7. https://doi.org/10.1145/3401956.3404187

Bird, S. (2020). Decolonising Speech and Language Technology. *Proceedings of the 28th International Conference on Computational Linguistics*, 3504–3519. https://doi.org/10.18653/v1/2020.coling-main.313

Blazquez, M. (2019). Using bigrams to detect written errors made by learners of Spanish as a foreign language. *CALL-EJ*, *20*, 55–69.

Blodgett, S.L., Barocas, S., Daumé, H., & Wallach, H. (2020). Language (technology) is power: A critical survey of" bias. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 454–476.

Bowker, L. (2020). Machine translation literacy instruction for international business students and business English instructors. *Journal of Business & Finance Librarianship*, *25*(1–2), 25–43. https://doi.org/10.1080/08963568.2020.1794739

Boyd, A. (2010). EAGLE: an Error-Annotated Corpus of Beginning Learner German. *Proceedings of LREC*.

Bragg, D., Koller, O., Bellard, M., Berke, L., Boudrealt, P., Braffort, A., Caselli, N., Huenerfauth, M., Kacorri, H., Verhoef, T., Vogler, C., & Morris, M. R. (2019). Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective. *ASSETS 2019*. https://www.microsoft.com/en-us/research/publication/sign-language-recognition-generation-and-translation-an-interdisciplinary-perspective/

Brandtzaeg, P. B., & Følstad, A. (2018). Chatbots: Changing user needs and motivations. *Interactions*, *25*(5), 38–43.

Brandtzaeg, P. B., & Følstad, A. (2017). *Why people use chatbots*. 377–392.

Bregler, C., Covell, M., & Slaney, M. (1997). Video Rewrite: Driving Visual Speech with Audio. *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, 353–360. https://doi.org/10.1145/258734.258880

Brown, C., Chauhan, J., Grammenos, A., Han, J., Hasthanasombat, A., Spathis, D., Xia, T., Cicuta, P., & Mascolo, C. (2020). Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 3474–3484. https://doi.org/10.1145/3394486.3412865

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.

Bruyn, P. C. D. (2020). Predicting behavioral profiles of online extremists through linguistic use of social roles. *Behavioral Sciences of Terrorism and Political Aggression*, *0*(0), 1–25. https://doi.org/10.1080/19434472.2020.1775675

Buckley, P., & Doyle, E. (2016). Gamification and student motivation. *Interactive Learning Environments*, *24*(6), 1162–1175. https://doi.org/10.1080/10494820.2014.964263

Cahill, A. (2015). Parsing learner text: To shoehorn or not to shoehorn. *The 9th Linguistic Annotation Workshop Held in Conjuncion with NAACL 2015*, 144.

Camgöz, N. C., Hadfield, S., Koller, O., Ney, H., & Bowden, R. (2018). Neural sign language translation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7784–7793. https://openaccess.thecvf.com/content_cvpr_2018/papers/Camgoz_Neural_Sign_Language_CVPR_2018_paper.pdf

Capuano, N., Greco, L., Ritrovato, P., & Vento, M. (2020). Sentiment analysis for customer relationship management: An incremental learning approach. *Applied Intelligence*. https://doi.org/10.1007/s10489-020-01984-x

Carosia, A. E. de O., Coelho, G. P., & Silva, A. E. A. da. (2019). The Influence of Tweets and News on the Brazilian Stock Market through Sentiment Analysis. *Proceedings of the 25th Brazillian Symposium on Multimedia and the Web*, 385–392. https://doi.org/10.1145/3323503.3349564

Carosia, A. E. O., Coelho, G. P., & Silva, A. E. A. (2020). Analyzing the Brazilian Financial Market through Portuguese Sentiment Analysis in Social Media. *Applied Artificial Intelligence*, *34*(1), 1–19. https://doi.org/10.1080/08839514.2019.1673037

Carr, N. (2020). *The shallows: What the Internet is doing to our brains*. WW Norton & Company.

Carrier, L. M., Rosen, L. D., Cheever, N. A., & Lim, A. F. (2015). Causes, effects, and practicalities of everyday multitasking. *Developmental Review*, *35*, 64–78. https://doi.org/10.1016/j.dr.2014.12.005

Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., & Way, A. (2017). Is Neural Machine Translation the New State of the Art? *The Prague Bulletin of Mathematical Linguistics*, *108*(1), 109–120. https://doi.org/10.1515/

pralin-2017-0013

Cavanaugh, J. M., Giapponi, C. C., & Golden, T. D. (2016). Digital technology and student cognitive development: The neuroscience of the university classroom. *Journal of Management Education*, *40*(4), 374–397.

Cettolo, M., Girardi, C., & Federico, M. (2012). WIT3: Web inventory of transcribed and translated talks. *Proceedings of the 16th EAMT Conference, 28-30 May 2012, Trento, Italy*, 8.

Chapelle, C. A., & Sauro, S. (Eds.). (2017). *The Handbook of Technology and Second Language Teaching and Learning*. Wiley Blackwell,.

Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing Machine Learning in Health Care—Addressing Ethical Challenges. *N Engl J Med.*, *378*(11), 981–983.

Chen, Y.-L., & Hsu, C.-C. (2020). Self-regulated mobile game-based English learning in a virtual reality environment. *Computers & Education*, *154*, 103910. https://doi.org/10.1016/j.compedu.2020.103910

Christensen, L. B. (2009). RoboBraille – Braille Unlimited. *The Educator*, *XXI*(2), 32–37.

Chu, C., Dabre, R., & Kurohashi, S. (2017). An Empirical Comparison of Domain Adaptation Methods for Neural Machine Translation. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 385–391. https://doi.org/10.18653/v1/P17-2061

Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, *118*(9). https://doi.org/10.1073/pnas.2023301118

Cobb, M. (2020). *The Idea of the Brain: The Past and Future of Neuroscience*. Profile.

Cohn, M., Liang, K. H., Sarian, M., Zellou, G., & Yu, Z. (2021). Speech rate adjustments in conversations with an Amazon Alexa socialbot. *Frontiers in Communication*, *6*, 82.

Consultancy.uk. (2019). *AR and VR to add £1.4 trillion to economy by 2030*. https://www.consultancy.uk/news/23135/ar-and-vr-to-add-15-trillion-to-economy-by-2030

Costa-Jussà, M. R. (2019). An analysis of gender bias studies in natural language processing. *Nature Machine Intelligence*, *1*(11), 495–496. https://doi.org/10.1038/s42256-019-0105-5

Coulthard, M., & Johnson, A. (2007). *An Introduction to Forensic Linguistics: Language in Evidence*. Routledge.

Coursey, K., Pirzchalski, S., McMullen, M., Lindroth, G., & Furuushi, Y. (2019). Living with Harmony: A Personal Companion System by Realbotix™. In Y. Zhou & M. H. Fischer (Eds.), *AI Love You: Developments in Human-Robot Intimate Relationships* (pp. 77–95). Springer International Publishing. https://doi.org/10.1007/978-3-030-19734-6_4

Cruz, A. F., Rocha, G., Sousa-Silva, R., & Cardoso, H. L. (2019). Team Fernando-Pessa at SemEval-2019 Task 4: Back to Basics in Hyperpartisan News Detection. *12th International Workshop on Semantic Evaluation (SemEval 2019)*.

Dale, E. (1946). *Audio-visual methods in teaching*. Dryden Press.

Dale, R. (2019). Industry Watch Law and Word Order: NLP in Legal Tech. *Natural Language Engineering*, *25*(1), 211–217.

Damiano, L., & Dumouchel, P. (2018). Anthropomorphism in Human–Robot Co-evolution. *Frontiers in Psychology*, *9*, 468. https://doi.org/10.3389/fpsyg.2018.00468

De Bruyn, P. C. (2020). Predicting behavioral profiles of online extremists through linguistic use of social roles. *Behavioral Sciences of Terrorism and Political Aggression*, 1–25. https://doi.org/10.1080/19434472.2020.1775675

De Cicco, R., Silva, S. C., & Alparone, F. R. (2020). Millennials' attitude toward chatbots: An experimental study in a social relationship perspective. *International Journal of Retail & Distribution Management*, *48*(11), 1213–1233. https://doi.org/10.1108/IJRDM-12-2019-0406

Denecke, K. (2008). Using SentiWordNet for multilingual sentiment analysis. *2008 IEEE 24th International Conference on Data Engineering Workshop*, 507–512. https://doi.org/10.1109/ICDEW.2008.4498370

Desmedt, B. (1995). *Herr Kommissar: A conversation simulator for intermediate German* (pp. 153–174).

Despotovic, V., Walter, O., & Haeb-Umbach, R. (2018). Machine learning techniques for semantic analysis of dysarthric speech: An experimental study. *Speech Communication*, *99*, 242–251. https://doi.org/10.1016/j.specom.2018.04.005

Devillers, L., Kawahara, T., Moore, R. K., & Scheutz, M. (2020). Spoken Language Interaction with Virtual Agents and Robots (SLIVAR): Towards Effective and Ethical Interaction (Dagstuhl Seminar 20021). *Dagstuhl Reports*, *10*(1), 1–51. https://doi.org/10.4230/DagRep.10.1.1

Dexter, A. (2021). *Oculus will sell you a Quest 2 headset that doesn't need Facebook for an extra $500. Is that the price of privacy?* https://www.pcgamer.com/oculus-will-sell-you-a-quest-2-headset-that-doesnt-need-facebook-for-an-extra-dollar500/

Díaz-Negrillo, A., & Fernández-Domínguez, J. (2006). Error Tagging Systems for Learner Corpora. *Revista Española de Lingüística Aplicada, ISSN 0213-2028, Vol. 19, 2006, Pags. 83-102*, *19*.

Dickinson, M., & Ragheb, M. (2015). *On Grammaticality in the Syntactic Annotation of Learner Language*. 158–167. https://doi.org/10.3115/v1/W15-1619

Dippold, D., Lynden, J., Shrubsall, R., & Ingram, R. (2020). A turn to language: How interactional sociolinguistics informs the redesign of prompt:response chatbot turns. *Discourse, Context & Media*, *37*, 100432. https://doi.org/10.1016/j.dcm.2020.100432

Dryer, M. S., & Haspelmath, M. (Eds.). (2013). *WALS Online*. Max Planck Institute for Evolutionary Anthropology.

Dubiel, M., Halvey, M., & Oplustil, P. (2020). Persuasive Synthetic Speech: Voice Perception and User Behaviour. *Proceedings of the 2nd Conference on Conversational User Interfaces*, 1–9. https://doi.org/10.1145/3405755.3406120

Dupont, M., & Zufferey, S. (2017). Methodological issues in the use of directional parallel corpora: A case study of English and French concessive connectives. *International Journal of Corpus Linguistics*, *22*. https://doi.org/10.1075/ijcl.22.2.05dup

Elfeky, M. G., Moreno, P., & Soto, V. (2018). Multi-Dialectical Languages Effect on Speech Recognition: Too Much Choice Can Hurt. *Procedia Computer Science*, *128*, 1–8. https://doi.org/10.1016/j.procs.2018.03.001

Emre, M. (2018). *The personality brokers: The strange history of Myers-Briggs and the birth of personality testing*. Doubleday.

Epstein, R. (2016). The empty brain [Website]. *Aeon*. https://aeon.co/essays/your-brain-does-not-process-information-and-it-is-not-a-computer

Erard, M. (2017). Why Sign-Language Gloves Don't Help Deaf People [Website]. *The Atlantic*. https://www.theatlantic.com/technology/archive/2017/11/why-sign-language-gloves-dont-help-deaf-people/545441/

Esuli, A., & Sebastiani, F. (2006). SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*. http://www.lrec-conf.org/proceedings/lrec2006/pdf/384_pdf.pdf

Fairclough, N. (1989). *Language and power*. Longman.

Felice, M. (2016). *Artificial error generation for translation-based grammatical error correction*.

Feng, X., Liu, M., Liu, J., Qin, B., Sun, Y., & Liu, T. (2018). Topic-to-Essay Generation with Neural Networks. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)*, 7.

Filhol, M., Hadjadj, M. N., & Testu, B. (2016). A rule triggering system for automatic text-to-sign translation. *Univ Access Inf Soc*, *15*, 487–498. https://doi.org/10.1007/s10209-015-0413-4

Finegan, E. (2008). *Language: Its Structure and Use* (6th, Inter ed.). Wadsworth.

Følstad, A., & Brandtzaeg, P. B. (2020). Users' experiences with chatbots: Findings from a questionnaire study. *Quality and User Experience*, *5*(1), 3. https://doi.org/10.1007/s41233-020-00033-2

Fortune Business Insights. (2021). *Chatbot Market Size and Regional Forecast, 2020-2027*. https://www.fortunebusinessinsights.com/chatbot-market-104673

Fryer, L., & Carpenter, R. (2006). Emerging technologies—Bots as language learning tools. *Language Learning & Technology*, *3*(10), 8–14.

Gafni, G., Thies, J., Zollhöfer, M., & Nießner, M. (2020). *Dynamic Neural Radiance Fields for Monocular 4D Facial Avatar Reconstruction*.

Gálvez, R. H., Beňuš, Š., Gravano, A., & Trnka, M. (2017). Prosodic Facilitation and Interference While Judging on the Veracity of Synthesized Statements. *INTERSPEECH*.

Gaucher, D., Friesen, J., & Kay, A.C. (2011). Evidence That Gendered Wording in Job Advertisements Exists and Sustains Gender Inequality. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/a0022530

Greenfield, P. M. (2009). Technology and informal education: What is taught, what is learned. *Science (New York, N.Y.)*, *323*(5910), 69–71. https://doi.org/10.1126/science.1167190

Grieve, J., Chiang, E., Clarke, I., Gideon, H., Heini, A., Nini, A., & Waibel, E. (2018). Attributing the Bixby Letter using n-gram tracing. *Digital Scholarship in the Humanities*, *April 2018*, 1–20.

Grosz, B., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, *12*(3), 175–204.

Guerra, A. M., Ferro, R., & Castañeda, M. A. (2018). Analysis on the gamification and implementation of Leap Motion Controller in the I.E.D. Técnico industrial de Tocancipá. *Interactive Technology and Smart Education*, *15*(2), 155–164. https://doi.org/10.1108/ITSE-12-2017-0069

Guenther, L., Ruhrmann, G., Bischoff, J., Penzel, T., & Weber, A. (2020). Strategic Framing and Social Media Engagement: Analyzing Memes Posted by the German Identitarian Movement on Facebook. *Social Media + Society*, *6*. https://doi.org/10.1177/2056305119898777

Hadjadj, M. N., Filhol, M., & Braffort, A. (2018). Modeling French Sign Language: A proposal for a semantically compositional system. *Proceedings of the Language Resource and Evaluation Conference (LREC)*.

Hansen, T., & Petersen, A. C. (2012). 'The Hunt for Harald'-Learning Language and Culture Through Gaming. *Proceedings of the 6th European Conference on Games Based Learning: ECGBL*, 184.

Hassani, K., Nahvi, A., & Ahmadi, A. (2016). Design and implementation of an intelligent virtual environment for improving speaking and listening skills. *Interactive Learning Environments*, *24*(1), 252–271. https://doi.org/10.1080/10494820.2013.846265

Hayles, K. (2012). *How We Think: Digital Media and Contemporary Technogenesis*. University of Chicago Press.

He, M., Xiong, B., & Xia, K. (2021). Are You Looking at Me? Eye Gazing in Web Video Conferences. *CPEN 541 HIT'21, Vancouver, BC, Canada*, 8.

Heift, T. (2002). Learner Control and Error Correction in ICALL: Browsers, Peekers, and Adamants. *CALICO Journal*, *19*.

Heift, T. (2003). Multiple Learner Errors and Meaningful Feedback: A Challenge for ICALL Systems. *CALICO Journal*, *20*(3), 533–548.

Herazo, J. (2020a). *Reconocimiento de señas de la lengua de señas panameña mediante aprendizaje profundo* [Masters Thesis, Universidad Carlos III de Madrid]. https://github.com/joseherazo04/SLR-CNN/blob/master/Jos%C3%A9%20Herazo%20TFM.pdf

Herazo, J. (2020b). Sign language recognition using deep learning. https://towardsdatascience.com/sign-language-recognition-using-deep-learning-6549268c60bd

Hewitt, J., & Kriz, R. (2018). Sequence-to-sequence Models [Lecture]. CIS 530 Computational Linguistics, U. Penn. https://nlp.stanford.edu/~johnhew/public/14-seq2seq.pdf

Hildt, E. (2019). Multi-Person Brain-To-Brain Interfaces: Ethical Issues. Frontiers in Neuroscience, 13, 1177. https://doi.org/10.3389/fnins.2019.01177

Hoang, H., Dwojak, T., Krislauks, R., Torregrosa, D., & Heafield, K. (2018). Fast Neural Machine Translation Implementation. 116–121. https://doi.org/10.18653/v1/W18-2714

Hodel, L., Formanowicz, M., Sczesny, S., Valdrová, J., & Stockhausen, L. von. (2017). Gender-Fair Language in Job Advertisements: A Cross-Linguistic and Cross-Cultural Analysis. Journal of Cross-Cultural Psychology, 48(3), 384–401. https://doi.org/10.1177/0022022116688085

Hoehn, S. (2019). Artificial Companion for Second Language Conversation. Springer International Publishing.

Hoehn, S., & Bongard-Blanchy, K. (2020). Heuristic Evaluation of COVID-19 Chatbots. Proceedings of CONVERSATIONS 2020.

Hoehn, Sviatlana. (2019). Artificial Companion for Second Language Conversation: Chatbots Support Practice Using Conversation Analysis. Springer International Publishing.

Holone, H. (2016). The filter bubble and its effect on online personal health information. Croatian Medical Journal, 57, 298–301. https://doi.org/10.3325/cmj.2016.57.298

Hovy, D., Spruit, S., Mitchell, M., Bender, E. M., Strube, M., & Wallach, H. (Eds.). (2017). Proceedings of the First ACL Workshop on Ethics in Natural Language Processing. Association for Computational Linguistics. https://doi.org/10.18653/v1/W17-16

Hutson, M. (2021). Robo-writers: The rise and risks of language-generating AI. Nature, 591, 22–25.

Ijaz, K., Bogdanovych, A., & Trescak, T. (2017). Virtual worlds vs books and videos in history education. Interactive Learning Environments, 25(7), 904–929. https://doi.org/10.1080/10494820.2016.1225099

inVentiv Health Communications. (2017). 2017 Digital Trends. https://www.gsw-w.com/2017Trends/inV_GSW_2017_Digital_Trends.pdf

Jain, S., Thiagarajan, B., Shi, Z., Clabaugh, C., & Matari□ , M. J. (2020). Modeling engagement in long-term, in-home socially assistive robot interventions for children with autism spectrum disorders. Science Robotics, 5(39). https://doi.org/10.1126/scirobotics.aaz3791

Jantunen, T., Rousi, R., Rainò, P., Turunen, M., Moeen Valipoor, M., & García, N. (2021). Is There Any Hope for Developing Automated Translation Technology for Sign Languages? In M. Hämäläinen, N. Partanen, & K. Alnajjar (Eds.), Multilingual Facilitation (p. 61□ 73).

Jia, J. (2009). CSIEC: A computer assisted English learning chatbot based on textual knowledge and reasoning. Knowledge-Based Systems, 22(4), 249–255. https://doi.org/10.1016/j.knosys.2008.09.001

Jiang, L., Stocco, A., Losey, D. M., Abernethy, J. A., Prat, C. S., & Rao, R. P. N. (2019). BrainNet: A Multi-Person Brain-to-Brain Interface for Direct Collaboration Between Brains. Scientific Reports, 9(1), 6115. https://doi.org/10.1038/s41598-019-41895-7

Jones, D. (2020). Macsen: A Voice Assistant for Speakers of a Lesser Resourced Language. Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-Resourced Languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL), 194–201. https://www.aclweb.org/anthology/2020.sltu-1.27

Juan, E. S. (2007). Language and Decolonization. In U.S. Imperialism and Revolution in the Philippines (pp. 67–87). Palgrave Macmillan US. https://doi.org/10.1057/9780230607033_4

Juola, P. (2015). The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions. Digital Scholarship in the Humanities, 30(suppl_1), i100–i113. https://doi.org/10.1093/llc/fqv040

Jurafsky, D. (2004). Pragmatics and computational linguistics. In L. R. Horn & G. Ward (Eds.), *Handbook of Pragmatics* (pp. 578–604). Blackwell Publishing.

Kasilingam, D. L. (2020). Understanding the attitude and intention to use smartphone chatbots for shopping. *Technology in Society*, *62*, 101280.

Kasper, G., & Wagner, J. (2011). A conversation-analytic approach to second language acquisition. In D. Atkinson (Ed.), *Alternative approaches to second language acquisition* (pp. 117–142). Taylor & Francis. https://doi.org/10.4324/9780203830932

Kastrati, Z., Imran, A. S., & Kurti, A. (2020). Weakly Supervised Framework for Aspect-Based Sentiment Analysis on Students' Reviews of MOOCs. *IEEE Access*, *8*, 106799–106810. https://doi.org/10.1109/ACCESS.2020.3000739

Khayrallah, H., & Koehn, P. (2018). On the Impact of Various Types of Noise on Neural Machine Translation. Proceedings of the 2nd Workshop on Neural Machine Translation and Generation, 74–83. https://doi.org/10.18653/v1/W18-2709

Klüwer, T. (2011). From Chatbots to Dialog Systems. In A. Sagae, W. L. Johnson, & A. Valente (Eds.), *Conversational Agents and Natural Language Interaction: Techniques and Effective Practices* (pp. 1–22). IGI Global. https://doi.org/10.4018/978-1-60960-617-6.ch016

Krummes, C., & Ensslin, A. (2014). What's Hard in German? WHiG: a British learner corpus of German. *Corpora*, *9*(2), 191–205.

Laguarta, J., Hueto, F., & Subirana, B. (2020). COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings. *IEEE Open Journal of Engineering in Medicine and Biology*, *1*, 275–281. https://doi.org/10.1109/OJEMB.2020.3026928

Lee, A., Prasad, R., Webber, B., & Joshi, A. (2016). Annotating discourse relations with the PDTB annotator. *Proceedings of the 26th International Conference on Computational Linguistics (COLING 2016): System Demonstrations*, 121–125.

Lee, J., Lee, J., & Lee, D. (2021). Cheerful encouragement or careful listening: The dynamics of robot etiquette at Children's different developmental stages. *Computers in Human Behavior*, *118*, 106697. https://doi.org/10.1016/j.chb.2021.106697

Lefer, M.-A., & Grabar, N. (2015). Super-creative and over-bureaucratic: A cross-genre corpus-based study on the use and translation of evaluative prefixation in TED talks and EU parliamentary debates. *Across Languages and Cultures*, *16*, 187–208. https://doi.org/10.1556/084.2015.16.2.3

Legault, J., Zhao, J., Chi, Y.-A., Chen, W., Klippel, A., & Li, P. (2019). Immersive Virtual Reality as an Effective Tool for Second Language Vocabulary Learning. *Languages*, *4*(1). https://doi.org/10.3390/languages4010013

Leviathan, Y., & Matias, Y. (2018). *Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone*. https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html

Li, B., Sainath, T. N., Sim, K. C., Bacchiani, M., Weinstein, E., Nguyen, P., Chen, Z., Wu, Y., & Rao, K. (2017). *Multi-Dialect Speech Recognition With A Single Sequence-To-Sequence Model*.

Li, M., Hickman, L., Tay, L., Ungar, L., & Guntuku, S. C. (2020). Studying Politeness across Cultures Using English Twitter and Mandarin Weibo. *Proc. ACM Hum.-Comput. Interact.*, *4*(CSCW2). https://doi.org/10.1145/3415190

Lilt Labs. (2017). *2017 Machine Translation Quality Evaluation*. https://web.archive.org/web/20170724060649/http://labs.lilt.com/2017/01/10/mt-quality-evaluation/

Lin, T.-J., & Lan, Y.-J. (2015). Language Learning in Virtual Reality Environments: Past, Present, and Future. *Journal of Educational Technology & Society*, *18*(4), 486–497.

Lo, S. L., Cambria, E., Chiong, R., & Cornforth, D. (2017). Multilingual sentiment analysis: From formal to informal and scarce resource languages. *Artificial Intelligence Review*, *48*(4), 499–527. https://doi.org/10.1007/s10462-016-9508-4

Lum, K., & Isaac, W. (2016). To predict and serve? *Significance*, *13*(5), 14–19. https://doi.org/10.1111/j.1740-9713.2016.00960.x

Macedo, D. (Ed.). (2019). *Decolonizing Foreign Language Education: The Misteaching of English and Other Colonial Languages*. Routledge.

Makransky, G., Terkildsen, T., & Mayer, R. (2017). Adding Immersive Virtual Reality to a Science Lab Simulation Causes More Presence But Less Learning. *Learning and Instruction*, *60*. https://doi.org/10.1016/j.learninstruc.2017.12.007

Marincat, N. (2020). Why the brain is not like a computer—And artificial intelligence is not likely to surpass human intelligence any time soon. *Medium.Com*. https://medium.com/is-consciousness/6d93d45df077

Markee, N. (2000). *Conversation analysis*. Routledge.

Matusov, E., Wilken, P., & Georgakopoulou, Y. (2019). Customizing Neural Machine Translation for Subtitling. *Proceedings of the Fourth Conference on Machine Translation (Volume 1: Research Papers)*, 82–93. https://doi.org/10.18653/v1/W19-5209

McEnery, T. (1995). *Computational Pragmatics: Probability, Deeming and Uncertain References* [Unpublished PhD thesis]. Lancaster University.

McLuhan, M., & Fiore, Q. (1967). *The medium is the message*. 123.

McLuhan, Marshall. (1962). *The Gutenberg galaxy. The making of typographic man*. University of Toronto Press.

McTear, M. (2020). Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots. *Synthesis Lectures on Human Language Technologies*, *13*(3), 1–251.

Meunier, F. (2020). A case for constructive alignment in DDL: Rethinking outcomes, practices and assessment in (data-driven) language learning. In P. Crosthwaite (Ed.), *Data-Driven Learning for the Next Generation. Corpora and DDL for Pre-tertiary Learners*. Routledge.

Naderi, N., & Hirst, G. (2018). Using context to identify the language of face-saving. *Proceedings of the 5th Workshop on Argument Mining*, 111–120. https://www.aclweb.org/anthology/W18-5214.pdf

Nagata, N. (2009). Robo-Sensei's NLP-Based Error Detection and Feedback Generation Robo-Sensei's NLP-Based Error Detection and Feedback Generation. *CALICO*, *26*.

Nangia, N., Vania, C., Rasika, B., & Bowman, S. R. (2020). CrowS-Pairs: A Challenge Dataset for Measuring Social Biases in Masked Language Models. *Proc. of EMNLP 2020*.

Naseem, T., Barzilay, R., & Globerson, A. (2012). Selective sharing for multilingual dependency parsing. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, *Volume 1: Long Papers*, 629–637.

Nass, C. (2004). Etiquette equality: Exhibitions and expectations of computer politeness. *Communications of the ACM*, *47*(4), 35–37.

Nass, C., & Yen, C. (2010). *The man who lied to his laptop: What we can learn about ourselves from our machines*. Penguin.

Nerbonne, J. (2016). Data from Non-standard Varieties. In S. Dipper, F. Neubarth, & H. Zinsmeister (Eds.), *Proceedings of the 13th Conference on Natural Language Processing: (KONVENS 2016)* (Vol. 16, pp. 1–12). Bochumer Linguistische Arbeitsbereichte.

Nguyen, M., He, T., An, L., Alexander, D. C., Feng, J., & Yeo, B. T. T. (2020). Predicting Alzheimer's disease progression using deep recurrent neural networks. *NeuroImage*, *222*, 117203. https://doi.org/10.1016/j.neuroimage.2020.117203

Nguyen, T., & Chiang, D. (2017). Transfer Learning Across Low-Resource Related Languages For Neural Machine Translation. *IJCNLP8 Proceedings*.

Noffs, G., Perera, T., Kolbe, S. C., Shanahan, C. J., Boonstra, F. M. C., Evans, A., Butzkueven, H., van der Walt, A., & Vogel, A. P. (2018). What speech can tell us: A systematic review of dysarthria characteristics in Multiple Sclerosis. *Autoimmunity Reviews*, *17*(12), 1202–1209. https://doi.org/10.1016/j.autrev.2018.06.010

Norman, E., & Furnes, B. (2016). The relationship between metacognitive experiences and learning: Is there a difference between digital and non-digital study media? *Computers in Human Behavior*, *54*, 301–309. https://doi.org/10.1016/j.chb.2015.07.043

North, M. M., & North, S. M. (2018). The Sense of Presence Exploration in Virtual Reality Therapy. *J-Jucs*, *24*(2), 72–84.

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M.-E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., … Staab, S. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. *WIREs Data Mining and Knowledge Discovery*, *10*(3), e1356. https://doi.org/10.1002/widm.1356

Ong, W. J. (1982). *Orality and Literacy. The Technologizing of the Word*. Routledge.

Pareja-Lora, A. (2012). OntoLingAnnot's Ontologies: Facilitating Interoperable Linguistic Annotations (Up to the Pragmatic Level). In C. Chiarcos, S. Nordhoff, & S. Hellmann (Eds.), *Linked Data in Linguistics: Representing and Connecting Language Data and Language Metadata* (pp. 117–127). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-28249-2_12

Pareja-Lora, A. (2014). The Pragmatic Level of OntoLingAnnot's Ontologies and Their Use in Pragmatic Annotation for Language Teaching. In E. Bárcena, T. Read, & J. Arús (Eds.), *Languages for Specific Purposes in the Digital Era* (pp. 323–344). Springer International Publishing. https://doi.org/10.1007/978-3-319-02222-2_15

Pareja-Lora, A., & Aguado de Cea, G. (2010). Modelling Discourse-Related Terminology in OntoLingAnnot's Ontologies. *Proceedings of TKE 2010: Presenting Terminology and Knowledge Engineering Resources Online: Models and Challenges*, 549–575.

Pareja-Lora, A., Blume, M., Lust, B. C., & Chiarcos, C. (2020). *Development of Linguistic Linked Open Data Resources for Collaborative Data-Intensive Research in the Language Sciences*. The MIT Press.

Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin.

Parmaxi, A. (2020). Virtual reality in language learning: A systematic review and implications for research and practice. *Interactive Learning Environments*, *0*(0), 1–13. https://doi.org/10.1080/10494820.2020.1765392

Patil, M., Chaudhari, N., Bhavsar, R., & Pawar, B. (2020). A review on sentiment analysis in psychomedical diagnosis. *Open Journal of Psychiatry & Allied Sciences*, *11*, 80. https://doi.org/10.5958/2394-2061.2020.00025.7

Pearson, J., Hu, J., Branigan, H. P., Pickering, M. J., & Nass, C. I. (2006). Adaptive language behavior in HCI: how expectations and beliefs about a system affect users' word choice. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1177–1180.

Peixoto, B., Pinto, D., Krassmann, A., Melo, M., Cabral, L., & Bessa, M. (2019). Using Virtual Reality Tools for Teaching Foreign Languages. In Á. Rocha, H. Adeli, L. P. Reis, & S. Costanzo (Eds.), *New Knowledge in Information Systems and Technologies* (pp. 581–588). Springer International Publishing.

Pereira, A. L. D. (2003). Problemas actuais da gestão do direito de autor: Gestão individual e gestão colectiva do direito de autor e dos direitos conexos na sociedade da informação. In *Estudos em Homenagem ao Professor Doutor Jorge Ribeiro de Faria – Faculdade de Direito da Universidade do Porto* (pp. 17–37). Coimbra Editora.

Perlin, K. (2016). Future Reality: How Emerging Technologies Will Change Language Itself. *IEEE Computer Graphics and Applications*, *36*(3), 84–89.

Petersen, K. A. (2010). *Implicit corrective feedback in computer-guided interaction: Does mode matter?* [PhD Thesis]. Georgetown University.

Peymanfard, J., Mohammadi, M. R., Zeinali, H., & Mozayani, N. (2021). Lip reading using external viseme decoding. *CoRR*, *abs/2104.04784*. https://arxiv.org/abs/2104.04784

Pierce, J. R., & Carroll, J. (1966). *Language and Machines: Computers in Translation and Linguistics*.

Pinto, A. G., Cardoso, H. L., Duarte, I. M., Warrot, C. V., & Sousa-Silva, R. (2020). Biased Language Detection in Court Decisions. In C. Analide, P. Novais, D. Camacho, & H. Yin (Eds.), *IDEAL (2)* (Vol. 12490, pp. 402–410). Springer. http://dblp.uni-trier.de/db/conf/ideal/ideal2020-2.html#PintoCDWS20

Plank, B. (2016). *What to do about non-standard (or non-canonical) language in NLP*. 8.

Poncelas, A., Lohar, P., Way, A., & Hadley, J. (2020). *The Impact of Indirect Machine Translation on Sentiment Classification*.

Popel, M., Tomkova, M., Tomek, J., Kaiser, □ ., Uszkoreit, J., Bojar, O., & □ abokrtský, Z. (2020). Transforming machine translation: A deep learning system reaches news translation quality comparable to human professionals. *Nature Communications*, *11*(1), 4381. https://doi.org/10.1038/s41467-020-18073-9

PwC EU Services. (2019). *Architecture for public service chatbots*. European Commission.

Ramesh, B. P., Prasad, R., Miller, T., Harrington, B., & Yu, H. (2012). Automatic Discourse Connective Detection in Biomedical Text. *Journal of the American Medical Informatics Association (JAMIA)*, *19*(5), 800–808.

Repetto, C. (2014). The use of virtual reality for language investigation and learning. *Frontiers in Psychology*, *5*, 1280. https://doi.org/10.3389/fpsyg.2014.01280

Research and Markets. (2020). *Intelligent Virtual Assistant Market Size, Share & Trends Analysis Report by Product (Chatbot, Smart Speakers), by Technology, by Application (BFSI, Healthcare, Education), by Region, and Segment Forecasts, 2020-2027*. https://www.researchandmarkets.com/reports/3292589/

Reznicek, M., Lüdeling, A., & Hirschmann, H. (2013). Competing target hypotheses in the Falko corpus: A flexible multi-layer corpus architecture. In A. Díaz-Negrillo, N. Ballier, & P. Thompson (Eds.), *Automatic Treatment and Analysis of Learner Corpus Data* (pp. 101–123). John Benjamins.

Reznicek, M., Lüdeling, A., Krummes, C., Schwantuschke, F., Walter, M., Schmidt, K., Hirschmann, H., & Andreas, T. (2012). *Das Falko-Handbuch. Korpusaufbau und Annotationen* (2.01). Humboldt Universität zu Berlin.

Robin, J., Harrison, J. E., Kaufman, L. D., Rudzicz, F., Simpson, W., & Yancheva, M. (2020). Evaluation of Speech-Based Digital Biomarkers: Review and Recommendations. *Digit Biomark*, *4*(99–108). https://doi.org/10.1159/000510820

Sagae, A., Johnson, W. L. & Valente, A. (Eds.) (2011). *Conversational Agents and Natural Language Interaction: Techniques and Effective Practices* (pp. 1–22). IGI Global.

Saleiro, P., Rodolfa, K. T., & Ghani, R. (2020). Dealing with Bias and Fairness in Data Science Systems: A Practical Hands-on Tutorial. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 3513–3514. https://doi.org/10.1145/3394486.3406708

Sandusky, S. (2015). *Gamification in Education*.

Saunders, J., Hunt, P., & Hollywood, J. S. (2016). Predictions put into practice: A quasi-experimental evaluation of Chicago's predictive policing pilot. *Journal of Experimental Criminology*, *12*(3), 347–371. https://doi.org/10.1007/s11292-016-9272-0

Scheufele, D. A. (1999). Framing as a theory of media effects. *Journal of Communication*, *49*, 103–122.

Schneider, B. (2020). What is 'correct' language in digital society? From Gutenberg to Alexa Galaxy [Blog post]. *Digital Society Blog*. https://www.hiig.de/en/what-is-correct-language-in-digital-society-from-gutenberg-to-alexa-galaxy/

Schneider, B. (Forthcoming). Von Gutenberg zu Alexa – Posthumanistische Perspektiven auf Sprachideologie. In M. Schmidt-Jüngst (Ed.), *Mensch – Tier – Maschine*. Transcript Verlag.

Schneider, S., Baevski, A., Collobert, R., & Auli, M. (2019). *wav2vec: Unsupervised Pre-training for Speech Recognition*.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press. https://books.google.pt/books/about/Speech_Acts.html?id=t3_WhfknvF0C&redir_esc=y

Ségouat, J. (2010). *Modélisation de la coarticulation en Langue des Signes Française pour la diffusion automatique d'informations en gare ferroviaire à l'aide d'un signeur virtuel* [PhD Dissertation, Université Paris Sud]. https://tals.limsi.fr/docs/TheseJeremieSegouat.pdf

Shahid, A. H., & Singh, M. P. (2020). A deep learning approach for prediction of Parkinson's disease progression. *Biomedical Engineering Letters*, *10*(2), 227–239. https://doi.org/10.1007/s13534-020-00156-7

Shalunts, G., Backfried, G., & Commeignes, N. (2016). *The Impact of Machine Translation on Sentiment Analysis*.

Shawar, B. A., & Atwell, E. (2007). Chatbots: Are they really useful? *Ldv Forum*, *22*(1), 29–49.

Sinclair, A., McCurdy, K., Lucas, C. G., Lopez, A., & Gaševi , D. (2019). Tutorbot Corpus: Evidence of Human-Agent Verbal Alignment in Second Language Learner Dialogues. *Proceedings of the 12th International Conference on Educational Data Mining*, 414–419.

Sinclair, A., McCurdy, K., Gasevic, D., Lucas, C., & Lopez, A. (2019). *Tutorbot Corpus: Evidence of Human-Agent Verbal Alignment in Second Language Learner Dialogues*.

Singer, L. M., & Alexander, P. A. (2017). Reading Across Mediums: Effects of Reading Digital and Print Texts on Comprehension and Calibration. *The Journal of Experimental Education*, *85*(1), 155–172. https://doi.org/10.1080/00220973.2016.1143794

Singh, M. (2021). *WhatsApp details what will happen to users who don't agree to privacy changes*. TechCrunch. https://techcrunch.com/?p=2115500

Solak, E., & Erdem, G. (2015). A Content Analysis of Virtual Reality Studies in Foreign Language Education. *Participatory Educational Research*, *2*(5), 21–26.

Sorgini, F., Caliò, R., Carrozza, M. C., & Oddo, C. M. (2018). Haptic-assistive technologies for audition and vision sensory disabilities. *Disability and Rehabilitation: Assistive Technology*, *13*(4), 394–421. https://doi.org/10.1080/17483107.2017.1385100

Soria, C. (2017). *The digital language vitality scale: A model for assessing digital vitality of languages*. 100–100.

Sousa Silva, R., Laboreiro, G., Sarmento, L., Grant, T., Oliveira, E., & Maia, B. (2011). 'twazn me□ ! ;(' Automatic Authorship Analysis of Micro-Blogging Messages. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 6716 LNCS* (pp. 161–168). https://doi.org/10.1007/978-3-642-22327-3_16

Sousa-Silva, R. (2013). *Detecting plagiarism in the forensic linguistics turn* [PhD Thesis]. Aston University.

Sousa-Silva, R. (2014). Detecting translingual plagiarism and the backlash against translation plagiarists. *Language and Law / Linguagem e Direito*, *1*(1), 70— – 94.

Sousa-Silva, R. (2018). Computational Forensic Linguistics: An Overview of Computational Applications in Forensic Contexts. *Language and Law / Linguagem e Direito*, *5*(2), 118–143.

Sousa-Silva, R. (2019). *When news become a forensic issue (revisited): Fake News, Mis- and Disinformation, and other ethical breaches* [Conference presentation]. IV International Symposium on Communication Management - XESCOM, Porto.

Sousa-Silva, R. (2021). Plagiarism: Evidence based plagiarism detection in forensic contexts. In Malcolm Coulthard, A. May, & R. Sousa-Silva (Eds.), *The Routledge Handbook of Forensic Linguistics* (2nd ed., pp. 364–381). Routledge.

Statista. (2020). *Augmented reality (AR) market size worldwide in 2017, 2018 and 2025*. https://www.statista.com/statistics/897587/

Statista. (2021). *Number of digital voice assistants in use worldwide from 2019 to 2024 (in billions)*. https://www.statista.com/statistics/973815/

Stepin, I., Alonso, J. M., Catala, A., & Pereira-Fariña, M. (2021). A Survey of Contrastive and Counterfactual Explanation Generation Methods for Explainable Artificial Intelligence. *IEEE Access*, *9*, 11974–12001.

Stoll, S., Camgöz, N. C., Hadfield, S., & Bowden, R. (2018). Sign language production using neural machine translation and generative adversarial networks. *Proceedings of the 29th British Machine Vision Conference (BMVC 2018)*. http://bmvc2018.org/contents/papers/0906.pdf

Sun, S., Guzmán, F., & Specia, L. (2020). Are we Estimating or Guesstimating Translation Quality? *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 6262–6267. https://doi.org/10.18653/v1/2020.acl-main.558

Sun, T., Gaut, A., Tang, S., Huang, Y., ElSherief, M., Zhao, J., Mirza, D., Belding, E., Chang, K.-W., & Wang, W. Y. (2019). Mitigating Gender Bias in Natural Language Processing: Literature Review. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 1630–1640. https://doi.org/10.18653/v1/P19-1159

Symonenko, S., Shmeltser, E., Zaitseva, N., Osadchyi, V., & Osadcha, K. (2020). *Virtual reality in foreign language training at higher educational institutions*.

Tatman, R. (2017). Gender and Dialect Bias in YouTube's Automatic Captions. *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, 53–59. https://doi.org/10.18653/v1/W17-1606

Taylor, J., & Richmond, K. (2020). Enhancing Sequence-to-Sequence Text-to-Speech with Morphology. *Submitted to IEEE ICASSP*. http://homepages.inf.ed.ac.uk/s1649890/morph/Morphology_interspeech2020.pdf

Tetreault, J. R., & Chodorow, M. (2008). Native Judgments of Non-Native Usage: Experiments in Preposition Error Detection. *Proceedings of the Workshop on Human Judgements in Computational Linguistics*, 24–32.

Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., & Niessner, M. (2016). Face2Face: Real-Time Face Capture and Reenactment of RGB Videos. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Timpe-Laughlin, V., Evanini, K., Green, A., Blood, I., Dombi, J., & Ramanarayanan, V. (2017). *Designing interactive, automated dialogues for L2 pragmatics learning*. 116–125. https://doi.org/10.21437/SemDial.2017-13

Tomalin, M., Byrne, B., Concannon, S., Saunders, D., & Ullmann, S. (2021). The practical ethics of bias reduction in machine translation: Why domain adaptation is better than data debiasing. *Ethics and Information Technology*. https://doi.org/10.1007/s10676-021-09583-1

Tseng, W.-T., Liou, H.-J., & Chu, H.-C. (2020). Vocabulary learning in virtual environments: Learner autonomy

and collaboration. *System*, *88*, 102190. https://doi.org/10.1016/j.system.2019.102190

Vanmassenhove, E., Hardmeier, C., & Way, A. (2018). Getting Gender Right in Neural Machine Translation. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 3003–3008. https://doi.org/10.18653/v1/D18-1334

Vázquez, C., Xia, L., Aikawa, T., & Maes, P. (2018). Words in Motion: Kinesthetic Language Learning in Virtual Reality. *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*, 272–276. https://doi.org/10.1109/ICALT.2018.00069

Vijayakumar, B., Höhn, S., & Schommer, C. (2018). Quizbot: Exploring formative feedback with conversational interfaces. *International Conference on Technology Enhanced Assessment*, 102–120.

Vilares, D., Alonso, M. A., & Gómez-Rodríguez, C. (2017). Supervised sentiment analysis in multilingual environments. *Information Processing & Management*, *53*(3), 595–607. https://doi.org/10.1016/j.ipm.2017.01.004

Wang, Changhan, Tang, Y., Ma, X., Wu, A., Okhonko, D., & Pino, J. (2020a). Fairseq S2T: Fast Speech-to-Text Modeling with fairseq. *ArXiv E-Prints*, arXiv:2010.05171.

Wang, Chien-pang, Lan, Y.-J., Tseng, W.-T., Lin, Y.-T. R., & Gupta, K. C.-L. (2020b). On the effects of 3D virtual worlds in language learning – a meta-analysis. *Computer Assisted Language Learning*, *33*(8), 891–915. https://doi.org/10.1080/09588221.2019.1598444

Wang, R., Newton, S., & Lowe, R. (2015). Experiential Learning Styles in the Age of a Virtual Surrogate. *International Journal of Architectural Research: ArchNet-IJAR*, *9*, 93–110. https://doi.org/10.26687/archnet-ijar.v9i3.715

Webber, B., Egg, M., & Kordoni, V. (2012). Discourse Structure and Language Technology. *Natural Language Engineering*, *18*(4), 437–490.

Weisser, M. (2014). Speech act annotation. In K. Aijmer & C. Rühlemann (Eds.), *Corpus Pragmatics: A Handbook* (pp. 84–116). Cambridge University Press.

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45.

Williams, K. (1991). Decolonizing the Word: Language, Culture, and Self in the Works of Ngũgĩwa Thiong'o and Gabriel Okara. *Research in African Literatures*, *22*(4), 53–61.

Willoughby, L., Iwasaki, S., Bartlett, M., & Manns, H. (2018). Tactile sign languages. In J.-O. Östman & J. Verschueren (Eds.), *Handbook of Pragmatics* (pp. 239–258). John Benjamins Publishing Company. https://doi.org/10.1075/hop.21.tac1

Wilske, S. (2014). *Form and Meaning in Dialog-Based Computer-Assisted Language Learning* [PhD Thesis]. University of Saarland.

Woolls, D. (2021). Computational forensic linguistics: Computer-assisted document comparison. In M. Coulthard, A. May, & R. Sousa-Silva (Eds.), *The Routledge Handbook of Forensic Linguistics* (2nd ed.). Routledge.

Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: A review. *Artificial Intelligence Review*, *53*(6), 4335–4385. https://doi.org/10.1007/s10462-019-09794-5

Yang, L., Li, Y., Wang, J., & Sherratt, R. S. (2020). Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. *IEEE Access*, *8*, 23522–23530.

Yule, G. (1996). *Pragmatics*. Oxford University Press.

Zellou, G., & Cohn, M. (2020). *Top-down effect of apparent humanness on vocal alignment toward human and device interlocutors*. 2020 Cognitive Science Society Meeting.

Zhang, Z., Wu, S., Liu, S., Li, M., Zhou, M., & Xu, T. (2019). Regularizing neural machine translation by target-bidirectional agreement. *Proc. AAAI, 33*, 443–450.

Zhang, T., Huang, H., Feng, C., & Wei, X. (2020). Similarity-aware neural machine translation: Reducing human translator efforts by leveraging high-potential sentences with translation memory. *Neural Computing and Applications*, 1–13.

Zhou, Y. (2019). Preventive Strategies for Pedophilia and the Potential Role of Robots: Open Workshop Discussion. In Y. Zhou & M. H. Fischer (Eds.), *AI Love You: Developments in Human-Robot Intimate Relationships* (pp. 169–174). Springer International Publishing. https://doi.org/10.1007/978-3-030-19734-6_9

Zhou, Y., & Fischer, M. H. (Eds.). (2019). *AI Love You: Developments in Human-Robot Intimate Relationships*. Springer.

Zhou, Z., Chen, K., Li, X., Zhang, S., Wu, Y., Zhou, Y., Meng, K., Sun, C., He, Q., Fan, W., Fan, E., Lin, Z., Tan, X., Deng, W., Yang, J., & Chen, J. (2020). Sign-to-speech translation using machine-learning-assisted stretchable sensor arrays. *Nat Electron*, *3*, 571–578.